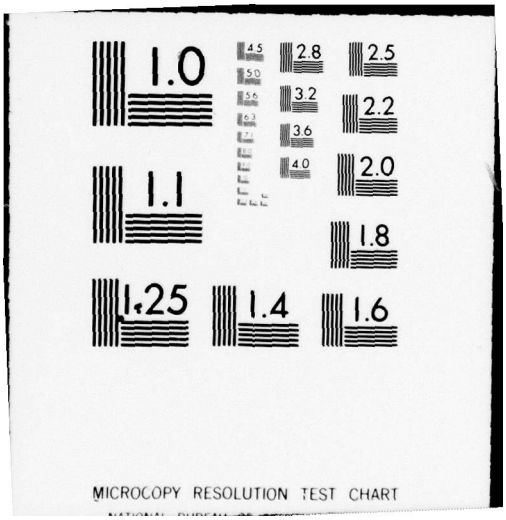
AD-A077 873 SCIENTIFIC SYSTEMS INC CAMBRIDGE MA F/G 4/2
MULTISPECTRAL CLOUD IDENTIFICATION STUDY.(U)
SEP 78 D E GUSTAFSON , W H LEDSHAM F19628-77-C-0208
UNCLASSIFIED AFGL-TR-78-0280 NL

1 OF 3
ADA
077873


1.0
1.1
1.25
1.4
1.6
1.8
2.0
2.2
2.5
2.8
3.2
3.6
4.0
4.5
5.0
5.6
6.3
7.1
MICROCOPY RESOLUTION TEST CHART

LEVEL II (10) B.S.

AFGL-TR-78-0280

MULTISPECTRAL CLOUD IDENTIFICATION STUDY

Donald E. Gustafson
William H. Ledsham
Mary G. Fowler
Eliot S. Blackman

Scientific Systems, Inc.
Suite 310
186 Alewife Brook Parkway
Cambridge, Massachusetts 02138

September 1978

Final Report
14 July 1977 to 14 August 1978

Approved for public release; distribution unlimited

AIR FORCE GEOPHYSICS LABORATORY
AIR FORCE SYSTEMS COMMAND
UNITED STATES AIR FORCE
HANSCOM AFB, MASSACHUSETTS 01731

Unclassified

SECURITY CLASSIFICATION OF THIS PAGE (When Data Entered)

(19) REPORT DOCUMENTATION PAGE	READ INSTRUCTIONS BEFORE COMPLETING FORM
1. REPORT NUMBER (18) AFGL-TR-78-0280	2. GOVT ACCESSION NO. / 3. RECIPIENT'S CATALOG NUMBER
4. TITLE (and Subtitle) (6) MULTISPECTRAL CLOUD IDENTIFICATION STUDY,	5. TYPE OF REPORT & PERIOD COVERED (9) Final Report, 14 Jul 77 - 14 Aug 78, / 6. PERFORMING ORG. REPORT NUMBER
7. AUTHOR(s) (10) Donald E. Gustafson, William H. Ledsham, Mary G. Fowler* Eliot S. Blackman†	8. CONTRACT OR GRANT NUMBER(s) (15) F19628-77-C-0208
9. PERFORMING ORGANIZATION NAME AND ADDRESS Scientific Systems, Inc. 186 Alewife Brook Parkway Cambridge, Massachusetts 02138	10. PROGRAM ELEMENT, PROJECT, TASK AREA & WORK UNIT NUMBERS 62101F (16) 669802AK (17) 03
11. CONTROLLING OFFICE NAME AND ADDRESS Air Force Geophysics Laboratory Hanscom AFB, Massachusetts 01731 Monitor/James Bunting/LYU	(11) 12. REPORT DATE Sep 78 / NUMBER OF PAGES
14. MONITORING AGENCY NAME & ADDRESS (if different from Controlling Office) (12) 278 393 341	15. SECURITY CLASS. (of this report) Unclassified / 15a. DECLASSIFICATION/DOWNGRADING SCHEDULE

16. DISTRIBUTION STATEMENT (of this Report)

Approved for public release; distribution unlimited.

17. DISTRIBUTION STATEMENT (of the abstract entered in Block 20, if different from Report)

18. SUPPLEMENTARY NOTES

*Environmental Research & Technology, Inc.
696 Virginia Road
Concord, MA 01742

†Bolt, Beranek & Newman, Inc.
50 Moulton Street
Cambridge, MA 02138

19. KEY WORDS (Continue on reverse side if necessary and identify by block number)

Satellite
Remote Sounding
Infrared Radiance
Microwave Radiance
Kalman Filters
Decision Trees

20. ABSTRACT (Continue on reverse side if necessary and identify by block number)

Both tactical and meteorological considerations demand a knowledge of cloud cover on a global basis. Due to the fact that there exist areas which are either data poor or denied by hostile forces, it is reasonable to infer cloud parameters in these regions from satellite-borne passive sensors. Both channel selection and inversion methodology to maximize the amount of cloud information are currently open questions.

(since areas exist)

(over)

DD FORM 1 JAN 73 1473 EDITION OF 1 NOV 65 IS OBSOLETE

Unclassified

SECURITY CLASSIFICATION OF THIS PAGE (When Data Entered)

393 342

This study deals with the problem of inferring both cloud and 'nuisance' parameters given four frequencies in the microwave region (10 GHz, 19 GHz, 37 GHz and 94 GHz) and ten frequencies in the thermal IR, near IR and visible region (.55 μ, .72 μ, 1.0 μ, 1.6 μ, 2.1 μ, 3.8 μ, 6.7 μ, 10.5 μ, 11.5 μ, and 12.5 μ). Using simulated data, the following studies were made: 1) Development and comparison of linear regression, nonlinear regression, extended Kalman filter and iterated extended Kalman filter inverters for cloud parameters using microwave data. The effect of channel dropouts was studied and a comparison with the Cramer-Rao bound was performed; 2) Development of nonlinear forward and inverse models in the visible through thermal IR region; 3) The development of a linear inverter using all frequencies; and 4) The development of nonparametric decision trees using visible through thermal IR to discriminate clear columns over loam, clear columns over snow, water clouds and ice clouds.

Results indicate that the iterated extended Kalman filter is a robust method of cloud parameter estimation, that while some parameters such as cloud top height and liquid water content are well estimated, others such as cloud thickness are not. Finally, simple decision trees may be formulated to distinguish the scenes analyzed.

TABLE OF CONTENTS

I. INTRODUCTION

The overall purpose of this study has been the development of methods to improve the identification of cloud features on a global basis through the interpretation of multispectral measurements obtained from satellite-borne passive spectrometers. For many data sparse areas, surface observations of cloud conditions are unavailable and cloud features can only be determined from satellite sensors. Inherent limitations in surface observations, such as the inability to detect clouds above an overcast layer, may be also resolved from space. Finally, requirements for data such as cloud heights, areal extent, cloud phase, rainfall and integrated water content makes desirable the use of all available information provided by surface and satellite observations. The specific parameters of interest include:

1) Cloud State
   - A) Glaciated cloud top and/or presence of glaciated cloud
   - B) Total cloud water content
   - C) Presence of rain and/or rainfall rate
   - D) Cloud height/thickness

2) Surface State
   - A) Foam coverage/wind speed over ocean
   - B) Snow/ice coverage

3) Integrated Atmospheric Water Vapor

A specific objective of this study has been to develop techniques which are suitable for inclusion in the operational Air Force 3-D Nephanalysis Program (3DNEPH), in terms of its cloud identification capability. A unifying statistical approach called Bayesian Decision Analysis is presented and its potential for application to multispectral cloud identification is shown.

1.1 Air Force Motivation

The Air Force at present lacks a capability for obtaining detailed specialized weather intelligence over many areas of the world. This intelligence is specially needed over denied areas of the battlefield and the data are also required by the Air Force Global Weather Central (AFGWC) in order to provide adequate meteorological support to many customers.

Experience in conflicts such as Vietnam has demonstrated the need for detailed cloud ceiling and visibility information over a target. The advantages of precision-guided munitions over conventional bombing dictate that precision-guided munitions be used whenever possible. Precision-guided munitions, however, require high cloud ceilings and good visibility to ensure that the pilots will be able to identify the target. Radiometric satellite sensors are already providing some cloud information, but proper processing is still badly needed in order to obtain better definition of cloud layering, cloud cover and cloud top heights over tactical-sized targets.

Due to the great climatic extremes that may be encountered in modern warfare, weather observations of specialized cloud parameters will be required under difficult observing conditions. The difference between successful and unsuccessful delivery of an intercontinental missile, for instance, may hinge on whether it is known that the reentry will take place through an all ice or an all water cloud. An adequate capability for processing presently existing and planned microwave sensor data will enhance the likelihood that such discrimination will be available when needed.

In order to achieve a solution to the total Air Force problem, effort will be required in two major areas. New and improved sensors must be developed in order to completely view those parts of the spectrum containing meteorologically significant information. The processing techniques for satellite data must be improved. Current techniques are often suboptimum in their retrieval capacity and the total volume of data taxes present processing resources.

1.2 Meteorological Satellite Sensors

1.2.1 Operational Satellite Systems

Present operational meteorological satellites include the DMSP, NOAA, and GOES systems. The satellites all carry visible and thermal-infrared radiometers. The two-channel imaging radiometers are designed to make observations of clouds and of the earth's surface. The visible channels are used to observe cloud distribution, cloud type, and snow

and ice distribution. The thermal infrared channels are used to measure cloud top temperatures and surface temperatures. Based on the measured thermal patterns, it is also possible to determine cloud distribution, high cloud types, and sea ice distribution; with corrections for atmospheric effects, sea surface temperatures can be measured directly using the thermal infrared.

The characteristics of each of the sensor systems are described in the following sections.

DMSP. The Department of Defense's Defense Meteorological Satellite Program (DMSP) consists of sun synchronous satellites in a polar orbit 450 nautical miles (830 km) above the earth. In the operational mode, two DMSP satellites provide imagery every six hours over any spot on earth (sunrise-sunset, noon-midnight). This imagery is in the visible/near infrared and thermal infrared spectral intervals over a 1600 nm (3200 km) swath below the satellite.

Real time meteorological data within the acquisition range, approximately a radius of 1400 nm (2800 km) of the receiving station, is provided to the DOD tactical sites. These sites are scattered around the globe and onboard U.S. Navy aircraft carriers. The Air Force Global Weather Central (AFGWC) at Offutt Air Force Base, Nebraska, receives stored data of global coverage from two command readout sites.

The present capability of the DMSP includes the following:

DATA TYPE	SPECTRAL INTERVAL	RESOLUTION
Light Fine (LF)	.4 to 1.1 μm	0.6 km
Thermal Fine (TF)	8 to 13 μm	0.6 km
Light Smoothed (LS)*	.4 to 1.1 μm	3 km
Thermal Smoothed (TS)	8 to 13 μm	3 km

*LS data has a low light capability that "sees" at night.

NOAA. The NOAA series of satellites carry the Scanning Radiometer (SR) and the Very High Resolution Radiometer (VHRR). Both the VHRR and SR data have been available twice-per-day from the polar-orbiting NOAA satellite (one daytime and one nighttime pass) since the first of the series was placed into operation in early 1973. The sensor characteristics are summarized below:

SENSOR	SPECTRAL INTERVAL	RESOLUTION
VHRR-Visible	0.5 - 0.7 μm	0.9 km
VHRR-Thermal Infrared	10.5-12.5 μm	0.9 km
SR-Visible	0.5 - 0.7 μm	4 km
SR-Thermal Infrared	10.5-12.5 μm	4 km

GOES. The third type of operational meteorological satellite data is from the GOES (Geostationary Operational Environmental Satellite). GOES-EAST launched in 1974, is stationed over the equator in an earth synchronous orbit in a position to view the eastern United States and Atlantic Ocean. GOES-WEST is in a position to view the western United States and eastern Pacific. Because of its earth synchronous orbit, GOES provides hemispheric coverage on an essentially continuous basis, with data being collected every one-half hour. Using the sequential coverage, cloud motions can be tracked, and the wind flow at cloud level can be derived. The characteristics of the Visual and Infrared Spin Scan Radiometer (VISSR) are as follows:

SENSOR	SPECTRAL INTERVAL	RESOLUTION
VISSR-Visible	0,55 - 0,75 μm	1 km
VISSR-Thermal Infrared	10.5 - 12.5 μm	8 km

1.2.2 Experimental Satellite Systems

The principal experimental meteorological satellites have been those of the NASA Nimbus series. These spacecrafts have carried many passive remote imagers including the THIR (Temperature Humidity Infrared Radiometer) infrared imager and the ESMR (Electrically Scanning Microwave Radiometer) microwave imager. The first Nimbus was launched in 1964; Nimbus-7 is scheduled for late 1978. The THIR has two channels, one in the 6.3-7.2 μm interval and one in the 10,5-12,5 μm window. The 6.7 μm channel has application for detecting atmospheric water vapor.

The initial microwave sensor was the ESMR (Electrically Scanning Microwave Radiometer) flown on Nimbus-5. The ESMR-5 has a frequency of 19.35 GHz and horizontal polarization whereas the ESMR-6 (Nimbus-6) has a frequency of 37.0 GHz and both horizontal and vertical polarizations; the maximum resolution of ESMR is of the order of 25 km. ESMR provides the capability of viewing the earth's surface through clouds and of mapping quantitative rainfall rates over oceans. Of the two frequencies, the 37.0 GHz instrument is more sensitive to detecting liquid clouds. The 37.0 GHz radiometer is also useful for detecting sea state and for mapping sea ice boundaries; it is also possible to distinguish new and old sea ice in the microwave data.

Although the ESMR-5 and ESMR-6 sensors have been single frequency radiometers, studies have indicated that a combination of frequencies, such as 19.35, 37.0, and 94.0 GHz, would likely provide considerably more information than either frequency alone. For example, clouds can be detected at 94 GHz, light rain at 37 GHz, and moderate rain at 19 GHz. By using a combination of frequencies, therefore, it should be possible to measure rainfall rates more accurately.

Although not designed for meteorological applications, Landsat and Skylab have also provided useful cloud information, both with regard to cloud structure and with regard to distinguishing cloud from the underlying surface. The Landsat Multispectral Scanner (MSS) has four channels in the visible and extending into the near-infrared; the spectral intervals are: 0.5-0.6 μm, 0.6-0.7 μm, 0.7-0.8 μm, and 0.8-1.1 μm.

The MSS has a resolution of 70-100 m. Although frequent repetitive coverage with Landsat is not possible, Landsat images have been found to be useful for studying detailed mesoscale cloud structure.

The Skylab missions were only for relatively short time periods, the longest being the 85 day Skylab-4 mission. Of particular interest, however, are the data from the 13-channel S192 Multispectral Scanner, which is the only spacecraft instrument that has provided measurements in several near-infrared bands out to 2 μm. Using the Skylab S192 data, studies have shown that the reflectance of snow surfaces drops dramatically in the near-infrared, such as in the spectral interval of 1.55-1.75 μm. The reflectance of water clouds, however, remains high in this same band. Therefore, it appears that an operational sensor in the near-infrared could be useful for distinguishing automatically between clouds and underlying snowcover. Moreover, aircraft measurements in the near-infrared have indicated the potential for distinguishing water droplet clouds from ice crystal clouds.

1.2.3 Summary of Applications of Present and Potential Sensors

The applications of sensors on present satellites and potential sensors are summarized below.

Visible	
0.4 - 1.0 μm	cloud distribution, cloud type, snow/ice distribution
0.55-0.90 μm	cloud distribution, cloud type, snow/ice distribution
0.725-1.0 μm	(red-near infrared) cloud distribution, cloud type, characteristics of snow/ice associated with melting
Near-IR	
1.6 μm	distinguish clouds from underlying snow/ice; distinguish water droplet clouds from ice crystal clouds
2.1 μm	distinguish clouds from underlying snow/ice; distinguish water droplet clouds from ice crystal clouds
3.8 μm	(near-IR/window) measure cloud-top and surface temperatures during nighttime
Thermal-IR	
6.7 μm	estimate atmospheric water vapor
8.0-13.0 μm	map cloud distributions, certain cloud types, ice distributions, estimate cloud temperatures
10.5-11.5 μm 11.5-12.5 μm	map cloud distributions, certain cloud types, ice distributions, estimate cloud top temperatures, sea surface temperatures

Microwave

10.7 GHz	estimate rainfall over ocean, detect sea state, determine snow and ice morphology
19 GHz	view earth's surface through clouds, detect sea state, measure rainfall over ocean, detect soil moisture, detect ice/water contrast
37 GHz	view earth's surface through some clouds, detect sea state, map ice boundaries, distinguish new and old ice, estimate light rainfall and some cloudiness over ocean, detect soil moisture
94 GHz	sense clouds over ocean, detect surface characteristics such as soil moisture, snow and ice, sea state, possibly detect rainfall over land

Visible and near-infrared sensors detect sunlight (or in some instances, moonlight) reflected from clouds and surfaces. Thermal-IR and microwave sensors detect emitted energy from clouds and surfaces. An exception is the 3.8 μm window which sees significant quantities of both solar and terrestrial energy during daytime.

1.3 Three Dimensional Nephanalysis

In addition to the generation of weather imagery, data from the Defense Meteorological Satellite Program satellites are input to a nephanalysis program (3DNEPH) which ultimately provides cloud height, coverage and type information to eventual users. Although the original documentation of the program (Coburn, 1971) implies a secondary role for

the satellite data with respect to its use in conjunction with surface, aircraft and other reports, the present operation considers the satellite data to be fundamental (Fye, 1978). Separate analyses are performed on data streams from the IR and visible spectral regions, to determine cloud coverage. Cloud heights are derived from the IR data alone, whereas cloud type estimates depend on visible, IR, or both forms of data depending on what is available.

One feature of analysis schemes used within the 3DNEPH program is the absence of explicit dependence on spatial information and reliance on computationally efficient procedures using first order statistics. Examples of these statistics are mean brightness and variability, defined as:

$$\text{mean brightness, } \overline{B} = \frac{1}{N} \sum_{i=1}^{N} B_i \qquad (1.3.1)$$

$$\text{and variability, } V = \frac{1}{2} + \frac{1}{N} \sum_{i=1}^{N} |B_i - \overline{B}| \qquad (1.3.2)$$

Because the data as used represent an 8 x 8 square array of 25 nm on a side, $N = 64$ and B_i represents data with a roughly 3 nm inter-sample interval. Infrared data are used to construct a histogram which generally is polymodal (usually one, two or three "distinct" modes). Cold modes are related to cloud layers. The warmest mode may be assigned to the ground.

In the simplest terms, the success of 3DNEPH has depended on the validity of two general rules and their converses:

1. Clouds have a high albedo (and the implied converse that a high albedo denotes a cloud); and
2. High clouds are colder than low clouds or ground (and the implied converse that a cold area denotes a high cloud).

However, these rules are not valid in a variety of special circumstances. Ambiguities arise, for example, when they are applied in estimating cloud amount over highly reflective coastlines, snow areas, deserts and mountains, and over cold regions such as the poles. The AFGWC has developed extensive processing to provide background brightnesses and temperatures. Bivariate analysis of visible and IR data resolves some ambiguities and is used for cloud typing. A good summary of the joint IR and visible characteristics of data from meteorological satellites is given in Anderson, et al. (1974).

The cloud field, described by the output variables of total cover, layered cloud amount, and cloud type, represents a mixture of classification data (cloud type), and estimation (total cover and layered amount). Data may be treated as strictly the output of a classification scheme by quantizing the formally continuous variables, and, in fact, present 3DNEPH procedures do produce reports of percentage cloud cover based on the number of cells in the 1/8 grid analysis window that are judged to be cloud covered. Within the classification schemes, decisions tend to be based on properties of samples as a group, e.g., average

brightness, modal (histogram) analysis, range, and variability. The total process is thus one of data reduction, where sensor output (sensed radiation) is subject to feature extraction which provides a relatively small number of important measures to the classification scheme. There are three major ways in which the above classification scheme can be aided. First, more and perhaps better input data could be provided by sensors which measure radiation in other frequency bands. (Note that the sensors often measure physical properties which only imply the output of the classifier, i.e., satellite radiometers directly measure energy and not cloud amount). Second, the list of derived statistics could be augmented, or improved in some other manner. Finally, action of the classifier could be changed by modifying the classification procedure.

A recent survey made by Pickett and Blackman (1976), summarizes the state of nephanalysis at AFGWC as of April, 1976. Before treating, in subsequent sections, some of the potential gains from joint use of information in several (electromagnetic) spectral bands, it is helpful to put the recent evolution of 3DNEPH in perspective.

Because of processing time constraints, 3DNEPH was originally designed to use little in the way of analytical power, and emphasized strong compression of the data stream. Furthermore, the compression was not simply on output, with the 64 visible and 64 IR radiometer data points processed to yield the 1/8 mesh output report (layers, cloud type and coverage), but involved intermediate compression or data reduction, as shown in Eqs. 1.3.1 - 1.3.2. In fact, the highest resolution data from the satellites is not used directly in 3DNEPH, again because of data

flow and processing time constraints. Given the scenario in which 3DNEPH was designed to operate, tradeoffs between computation time and full use of data (dimensionality of feature space) were reasonably made. With the advent of higher resolution and multispectral systems, however, the data compression problem is more severe. Yet, justification for these new systems requires that these new data be fully exploited.

The improvement of remote sensing capability primarily in terms of increased (electromagnetic) spectral information makes amenable to solution several problems (e.g., cloud identification) which could not previously be properly treated. Also, some of the simplifying assumptions implicit in 3DNEPH, e.g., that satellite radiometers sense cloud top temperature and therefore cloud top height rather than radiation both from the cloud and upwelling through the cloud, need not be made. On the other hand, assumptions of this type may well be acceptable for low priority areas. This dichotomy suggests strongly that upgrading 3DNEPH take two complementary paths: full use of all data in selected areas, to solve important problems (cloud amount, phase, ice vs. water), directly, and efficient feature extraction for all other cases.

In the satellite IR data processor, several different types of computation are performed. The difference between surface and radiometer-sensed temperature is used to infer presence of clouds (a yes or no type of decision) possibly yielding misinterpretation of low fog or stratus as a clear sky condition. Total cloud cover is computed using a decision matrix approach based on temperature differences. These tests may be sharpened using statistical measures of cloud presence and type.

In the current version of 3D Nephanalysis, infrared data is merged with other data in several ways. If video data is also available, then the total cloud cover estimate is based on the most recent measurement. If both types of measurement were obtained at the same time, then the total cloud cover is assumed to be the higher of the two computed values. Both of these procedures can be analyzed to determine if a statistical merging technique would yield more accurate results.

1.4 Previous Techniques for Extraction of Cloud Information

Numerous techniques for extracting cloud information from satellite data have been developed over the past several years. Examples of the information which may be obtained are cloud cover, cloud height and thickness, cloud type, phase (ice or water), layering and precipitation. In the remainder of this subsection some of the most significant techniques and results pertaining to the problem of obtaining cloud descriptions from satellite data are presented. All of these techniques are based on the assumption that temperature decreases monotonically with increasing altitude.

1.4.1 Cloud Cover

There are several techniques for estimating the amount of cloud cover in a particular region being scanned by passive sensors. One technique presently used in 3DNEPH is a frequency histogram of infrared observations. Cloud cover can be estimated from the probability density

distribution of the difference between earth surface temperature and the equivalent observed black-body temperature, using the 10.5-12.5 μm atmospheric "window" band. If it is assumed that clouds are of one type with a characteristic cloud-top temperature, the surface temperature is constant and the clouds are black with respect to IR flux, then the ideal frequency distribution is shown in Fig. 1.4.1.

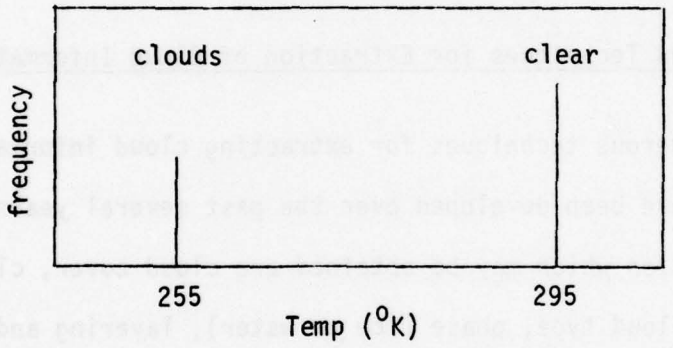


Figure 1.4.1. Ideal Frequency Distribution

If the observations are evenly distributed over the sensed region, then the height of each peak is proportional to the fractional cover of the thpe associated with the peak. A more realistic profile is given in Fig. 1.4.2.

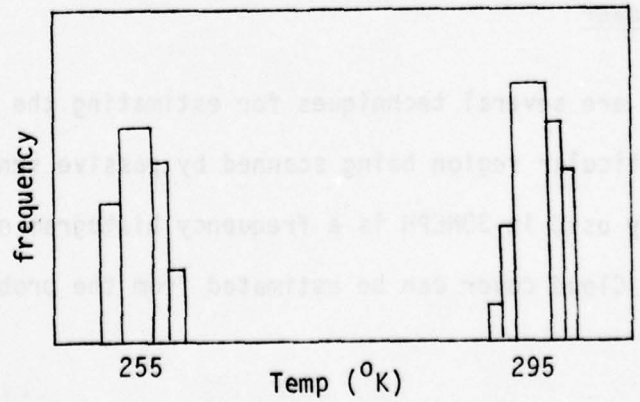


Figure 1.4.2. Realistic Frequency Distribution

Now the fractional cover of each type is proportional to the area under each mode. This technique can be extended to include two or more cloud types. NIMBUS II MRIR channel 2 data (10-11 μm) has been analyzed using this technique by Lo and Johnson (1971). This approach has also been used by Rao (1970).

In other studies, Fritz and Rao (1967) used measurements at 6 μm and 10 μm to locate regions of substantial cloudiness. Koffler, et al. (1973) used ITOS I data to obtain cloud amount. The data used was IR data at 11 μm and temperature measurements at 400 and 700 mb. It was found that there was a tendency to overestimate the total cloud amount.

Rao (1970) proposed a simple technique for cloud cover determination. Let TD = {surface temperature - radiation temperature from satellite data}. Then, approximately

$$TD = \frac{\partial T}{\partial Z} \Delta Z, \qquad (1.4.1)$$

where $-(\partial T/\partial Z)$ is the lapse rate, assumed constant, and ΔZ is the height difference between the radiating surface and the ground. Large values of TD indicate thick clouds at high and middle levels, while small values of TD indicate relatively cloud-free conditions. The use of a relative measure such as TD helps to eliminate the problem of cold surface temperatures (due to snow and ice) which give low absolute temperature readings.

One reason for determining the percentage of cloud cover is to simplify calculation of other meteorological parameters such as vertical temperature profile. The fundamental quantities of interest are the "clear-column radiances", which are the equivalent cloud-free radiances. Using these, the inverse problem is solved to recover the temperature profiles. The clear-column radiance at frequency ν may be estimated from (Smith, et al., 1974)

$$I_c(\nu) = \frac{I_1(\nu) - NI_2(\nu)}{1 - N},$$

where $I_1(\nu)$, $I_2(\nu)$ are spatially-independent measured radiances and

$$N = \frac{N_1}{N_2} = \frac{I_c(w) - I_1(w)}{I_c(w) - I_2(w)}$$

with $I_1(w)$, $I_2(w)$ selected to ensure that $0 \leq N \leq 1$, and where N_i is the fractional cloud cover in the region producing measurement $I_i(\nu)$. $I_c(w)$, the clear column radiance, is obtained by using the 10-11 μm atmospheric window for IR measurement. Thus, this technique can be used to infer the relative percentages of cloud cover in regions which have the same surface temperature but different observed equivalent black-body temperatures due to cloud cover differences.

1.4.2 Cloud Height and Thickness

A technique due to Rao (1970) uses TD to estimate approximate cloud height. Figure 1.4.3 depicts the qualitative relationship between frequency distribution and temperature for several typical cloud heights. Thus, an estimate of cloud height may be obtained by matching observed data to the curves of the figure.

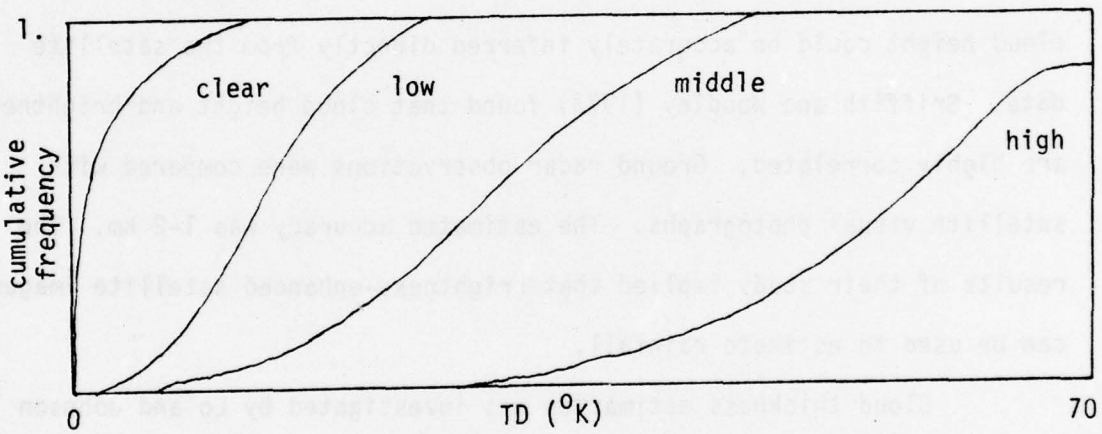


Figure 1.4.3. Hypothetical Cumulative Frequency Distribution of TD (^{o}K) for Various Cloud Heights

Since there is generally very little water vapor above the clouds, a radiometer, in the presence of thick clouds, will record the energy associated with the cloud top, which may be 40^{o}K or more below the surface temperature. Many factors influence the accuracy of the estimates, including cloud amount, surface temperature, and vertical distribution of water vapor and temperature.

In their study using TIROS I data, Koffler, et al. (1973) were able to specify "no clouds", "low clouds", "middle clouds" or "high clouds", using IR and 400 and 700 mb temperatures only. However, cirrus clouds were found difficult to locate and there was a tendency to underestimate cloud height. Reynolds and VonderHaar (1973) used visible data from the ATS-3 geosynchronous satellite to infer cumulus cloud height and growth. Cloud height derived from ground radar was compared to reflected solar radiance in the green region. By linearly fitting the data it was found that a correlation of 0.88 existed, implying that cloud height could be accurately inferred directly from the satellite data. Griffith and Woodley (1973) found that cloud height and brightness are highly correlated. Ground radar observations were compared with satellite visual photographs. The estimated accuracy was 1-2 km. The results of their study implied that brightness-enhanced satellite imagery can be used to estimate rainfall.

Cloud thickness estimation was investigated by Lo and Johnson (1971) who found that clouds with large vertical extent have low equivalent blackbody temperatures.

1.4.3 Cloud Type

Fritz and Rao (1967) found that cirrus clouds have a higher transmission for radiation at 10 μm than for 6 μm (about 50% versus almost opaque). In their study, Lo and Johnson (1971) found that clouds with high water vapor content (either in the liquid or frozen

state) have lowest equivalent blackbody temperatures. They studied data from NIMBUS II MRIR channels 1 (6.4-6.9 μm) and 2 (10-11 μm). The channel 2 temperature difference (TD2) was studied, with TD = {surface temperature - radiation temperature from satellite data}. High values of TD2 indicated regions of cumulonimbus activity. High values of the channel 1 temperature difference (TD1) were found to indicate cirrus cloudiness. Using both channels enabled the water vapor content above low-level clouds to be estimated.

Shenk, et al. (1976) used four channels of the NIMBUS-3 MRIR to identify unique individual cloud types and cloud type combinations over tropical oceans. The channels used were 0.2-4.0 μm, 6.5-7.0 μm, 10-11 μm and 20-23 μm and the cloud categories that were identified included cirrus, cirrostratus, stratus or stratocumulus, and cumulonimbus. A decision matrix is set up to determine both single cloud types and appropriate cloud combinations. The method may be extendable to include ocean areas in middle and high latitudes. Channel deletion was studied to determine sensitivity of the solutions to loss of data. Deletion of the 20-23 μm channel apparently had no adverse effect for tropical regions but may be important at middle latitudes. With just the 6.5-7.0 μm and 10-11 μm channels, there was a tendency to overestimate the cumulonimbus cover and middle cloud cover beneath a high cirrus overcast. With just the 6.5-7.0 μm channel areas with reasonably dense cirrus were mapped successfully.

Following Anderson, et al. (1974), 3DNEPH currently uses a bivariate distribution of IR and visible satellite data to infer cloud

types. Among the distinct cloud types which could be discriminated were thin cirrus, dense cirrus, stratus and stratocumulus, and cumulonimbus. Booth (1973) and Sikula (1974) have used information bearing on cloud size to help in cloud type classification. Power spectra of visible and IR data have been combined with first-order measures such as percentile brightness and temperature values, to provide "features" upon which a linear discriminant classifier acted.

1.5 Thrust of Study

The principal thrust of this study is the development of efficient and robust techniques for inference of meteorological parameters from multispectral satellite-borne spectrometers. The parameters under consideration include cloud parameters, atmospheric parameters and surface parameters, over both ocean and land.

It is argued that the solution involves optimum processing of information, in which available *a priori* information is combined with the information carried by the remotely-sensed data. Since both types of information carry uncertainty, the appropriate methodology is, of necessity, a statistical one.

A systematic and comprehensive approach to this problem is put forth which is based on Bayesian decision analysis. Specific techniques are developed to handle both continuous-valued and discrete-valued parameters. This approach offers several distinct advantages over deterministic methods or direct statistical approaches such as regression techniques:

(1) The physics of radiative transfer may be used directly through use of well-understood mathematical models.

(2) Sensor noise can be treated directly.

(3) The a priori information and the information carried by the available data are combined in a statistically optimum manner.

(4) Nonlinearities may be handled in a very direct manner.

(5) Missing data or degraded measurements are handled directly in a statistically optimum manner.

The approach to estimation of continuous-valued parameters is based on Kalman filtering theory. A particular advantage is that the theory is easily extendable to include the dynamic problem, in which parameters are updated in real time whenever measurements become available. This is possible since the Kalman filter carries along not only the best estimate of present values of all parameters, but also an estimate of the covariance matrix of estimation errors; that is, the Kalman filter is fully recursive.

The approach to inference of discrete-valued parameters involves the use of decision trees determined by nonparametric statistical methods. Their performance approaches, aysmptotically, the Bayes error rate.

With the Bayesian approach it is possible to compute absolute upper bounds of performance, even in the nonlinear case, for a given channel set as a function of the a priori parameter uncertainties and the sensor noise variances.

This approach is evaluated to assess the possibility of incorporation into the 3DNEPH program. The scope of the study has been necessarily restricted in several ways. First, the data base was generated exclusively from detailed simulations at ERT and AFGL in order to procure a sufficiently large validated data set. This allows quantitative evaluation of sensitivity to noise, missing data and other modeling errors and direct comparison of techniques. The study was restricted to one (vertical) dimension; i.e., the horizontal field was assumed to be absolutely homogeneous over the field of view of each sensor.

II. PHILOSOPHY OF APPROACH

In previous sections, various specialized techniques for cloud identification using satellite spectral data were discussed. These techniques have been developed, expanded, and improved as different types of data (visible, IR, microwave, etc.) have become available. Most of the techniques and algorithms were first developed to handle a single type of data. Some of these techniques have been augmented to include more than one type of sensor, but a unified approach to the problem is still lacking. In this section, we discuss a statistical methodology for combining information from different sources and for making optimum decisions. The basis of this methodology is Bayesian Decision Analysis. The reasons for considering statistical approaches are that observations are contaminated with noise, different measurements have different accuracies and models contain random errors.

2.1 Bayesian Decision Analysis

Statistical inference and decision analysis involve combining different sets of observations to compute probabilities of occurrence of different events. In the Multispectral Cloud Identification Study, the observations are represented by radiance measurements at different frequencies and the events are represented by different cloud types, heights, phases, mass and other parameters. It is possible to define a vector of parameters for a cloud such that the values taken by the

vector may be called a state vector of the cloud and will be denoted by x. Notice than an element of x corresponding to cloud height will be a continuous variable. Other elements of x would include cloud phase, thickness, amount and extent.

We will denote by y the vector of spectral measurements from the satellite during one pass. Generally, y would correspond to measurements over a single field of view since this is the easiest case to consider. In certain cases, better cloud identifications may be obtained by considering observations and/or their statistics based on a number of adjacent fields of view. It is assumed here that the raw observations have been processed before they are incorporated in the vector y.

The basic calculation of Bayesian Inference is to obtain the posterior probability distribution of x given y, denoted mathematically* by $p(x|y)$. The celebrated Bayes Formula gives this as

$$p(x|y) = \frac{p(y|x)\ p(x)}{\int_{\Omega_x} p(y|x)\ p(x)\ dx} \qquad (2.1.1)$$

Equation (2.1.1) performs elegantly the "inversion" process used in many applications including IR inversions to obtain temperature profiles. The probability $p(y|x)$ denotes the conditional probability of observing y given that x has occurred. For example, if x corresponds to a high cirrus cloud, then $p(y|x)$ is specified using theoretical and empirical radiative transfer models for cirrus clouds. Thus, the solution to the "direct" or "forward" problem along with a sensor noise model specifies $p(y|x)$. The probability $p(y|x)$ is defined for all values of x in the set Ω_x of relevant cloud parameter values.

---

*All marginal probability distributions are denoted by $p(\cdot)$ and conditional distributions as $p(\cdot|\cdot)$.

The probability p(x) denotes a priori probability of x before y is observed. It may be obtained using climatological data and measurements from previous passes. Global cloud models that have been developed over the last decade and certain known properties such as height and extent of different cloud types will be used to define the probability p(x). It should be pointed out that some of the temperature inversion techniques make use of prior data but do so only in a heuristic fashion.

Bayes formula, viz. Eq. (2.1.1), computes the probability of x given y, i.e. the probability of each event x given a particular observation y. In this sense, it inverts the "direct" probability p(y|x), which has to be defined over the space of all possible events denoted by Ω_x.

Once the posterior probabilities p(x|y) are obtained, a decision as to the most likely x may be obtained by simply picking the x value with the highest posterior probability. For certain other applications, the mean or the median of p(x|y) may be a more appropriate value to use.

Advantages of the Bayesian Decision Analysis Approach

The Bayes Decision Analysis approach is a comprehensive approach and includes specialized approaches such as Regression, Decision Trees and Minimum Information methods as special cases. In the statistical literature, it is recognized to be the best approach for combining information from different sources. It has been used extensively in system identification, hypothesis testing, communication systems, and navigation systems. For example, the navigation system of modern fighter aircraft combines inertial and sensor information using a Kalman Filter which is based on the Bayes Formula.

The Bayesian Decision Analysis approach is also flexible in the sense that new types of measurements can be added without changing the structure of the algorithm. For computational efficiency, it is customary to choose probabilities $p(y|x)$ and $p(x)$ from the so-called "reproducing conjugate" family which has the property that the prior and the posterior distributions are of the same type. For example, if $p(x)$ is chosen as multivariate normal and $p(y|x)$ is also multivariate normal, then $p(x|y)$ automatically turns out to be multivariate normal also. In this case, only the means and covariances have to be computed and the integration in the denominator of Eq. (2.1.1) is performed analytically.

2.2 Estimation of Continuous-Valued Parameters

Parameters such as cloud height, cloud thickness, integrated liquid water content, integrated water column, ocean surface wind speed and rain rate may be regarded as continuously variable between certain limits. These parameters may be estimated using the tools of modern estimation theory, in particular, the Kalman Filter and some of its variations. This technique is a way of estimating variables which change with time, as do all of the variables of interest in this study. However, in this study, we have assumed all variables to be static; nevertheless, the Kalman Filter provides an appropriate methodology for the static problem, and possesses the advantage of being extendable in a straightforward manner to the dynamic problem at a later time. For this reason, a discussion of the Kalman Filter, in its full dynamic context is now given.

2.2.1 Kalman Filtering Method [Kalman, 1960]

The Kalman Filter approach to estimation is based on certain assumptions made on the nature of the variables to be estimated, the measurement process and noises corrupting the system. The variables to be estimated are collected together in a state vector (x) which evolves as a Gauss-Markov process with known statistics. The measurements are constrained to be linearly related to the state and are corrupted by additive white Gaussian noise with known statistics. Under these assumptions, the conditional mean (minimum-variance estimate) of x, given all of the data, may be computed recursively. The error covariance matrix of the estimate is also computed recursively, but is independent of the values of the measurements. The conditional mean is Gaussianly distributed and thus all statistical information regarding the state is available.

The generality of the state variable formulation and the ease of filter implementation on digital computers has led to its great popularity, especially in aerospace problems. Multivariate estimation problems can be handled systematically for either time-varying or time-invariant systems. A wealth of experience has been gained over the past fifteen years in solving many diverse problems. These include spacecraft, marine and aircraft navigation, signal processing, industrial process control, economic forecasting, satellite orbit determination, and image processing.

The cloud identification problem is an example of one in which the assumptions of Gaussianness, linearity and known noise statistics do not strictly hold. In this case, extended Kalman filtering theory,

which is discussed in Section 2.4, is expected to produce better results than other techniques such as the statistical D-method, the statistical eigenvector method, or regression techniques. The reason for this is that the Kalman filter formulation allows a much more complete statistical description of the problem to be utilized. The Kalman filter formulation is based on a state-space representation of the system to be analyzed. We now briefly describe the state-space approach.

2.2.2 State Space Models

State space models of random processes are based on the Markov property which, in simple terms, implies the independence of the future of the process from its past given the present state. In other words, the state of a Markov process summarizes all of the information from the past that is nessary to predict its future. For obvious reasons, only the case where the state vector is finite-dimensional is of practical interest. A general state vector model is typically specified in terms of the following five quantities (see also Figure 2.2.1):

(i) three vectors, respectively, of input, output and internal state variables;

(ii) a rule for transformation of the state vector from one time instant to the next;

(iii) a relationship between the input, output and state variables;

(iv) initial state;

(v) joint statistics of all random variables.

Mathematically,

$$x(k+1) = f(x(k),k) + G(k)\, w(k) \qquad (2.2.1)$$

$$y(k+1) = h(x(k+1), k+1) + v(k) \qquad (2.2.2)$$

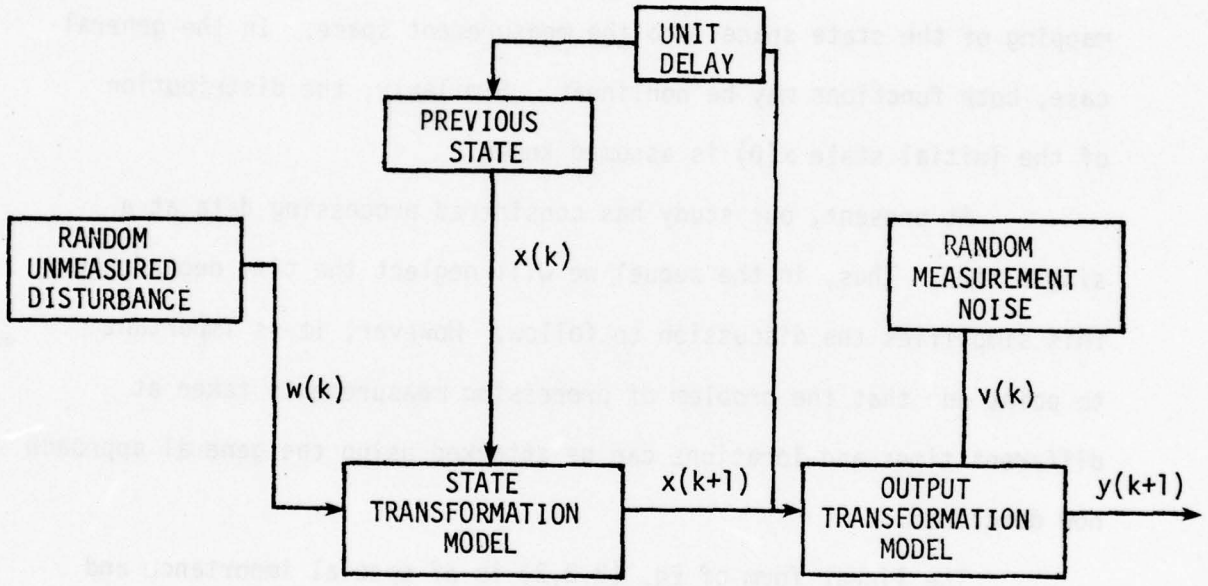


Figure 2.2.1 State Vector Model

where $x(k)$ is the $n \times 1$ state vector at time t_k, $w(k)$ is the $n \times 1$ process noise vector, $y(k)$ is the $m \times 1$ measurement vector, $v(k)$ is the $m \times 1$ measurement noise vector and $w(k)$ are assumed to be uncorrelated white noise sequences with known distributions. The function $f(x(k),k)$ represents the deterministic portion of the state transition from time k to k+1. The function $h(x(k+1),k+1)$ represents the noiseless mapping of the state space into the measurement space. In the general case, both functions may be nonlinear. Similarly, the distribution of the initial state $x(0)$ is assumed known.

At present, our study has considered processing data at a single time. Thus, in the sequel we will neglect the time dependence. This simplifies the discussion to follow. However, it is important to point out that the problem of processing measurements taken at different times and locations can be attacked using the general approach now described.

The linear form of Eq. (2.2.2) is of special importance and is written as

$$y = Hx + v, \tag{2.2.3}$$

If v and x are Gaussian then so is y. The state is assumed Gaussian with mean x_o and covariance matrix P_o. The noise $v(k)$ is assumed to have zero mean and covariance matrix V. The measurement matrix H is deterministic. The principal advantage of this representation is that the posterior distribution $p(x|y)$ of x, given y, turns out to be Gaussian and its first two moments are computed by the Kalman filter.

2.2.3 Kalman Filter Equations

Denote the mean and covariance of the Gaussian density function $p(x|y)$ by $\hat{x}$ and P, respectively. Then the Kalman Filter equations are given as follows:

$$\hat{x} = x_o + M\nu \tag{2.2.4}$$

$$\nu = y - H x_o \tag{2.2.5}$$

$$M = P_o H^T \sum^{-1} \tag{2.2.6}$$

$$\sum = H P_o H^T + V \tag{2.2.7}$$

$$P = P_o - M H P_o \tag{2.2.8}$$

where ν, the "innovations," represents the differences between the predicted and actual measurements. This quantity may be regarded as the new information brought in by the observation y. The quantity $\sum$ represents the covariance matrix of the innovations and M is a given matrix (often called the Kalman gain) that is used to combine the innovations with the prior estimate x_o to yield the estimate $\hat{x}$.

2.2.4 Properties of Kalman Filters

Kalman Filters possess a number of desirable properties which are of practical interest. These properties are discussed briefly here.

Optimality. The Kalman filter estimate $\hat{x}$ is the minimum variance estimate of x given the measurement y. In addition, since the mean, mode and median of the Gaussian density are identical, $\hat{x}$ is optimal for a whole class of symmetric loss functions. Since the Kalman Filter computes the first two moments which completely characterize a Gaussian random variable, all statistical information regarding x contained in the measurement y is available from the Kalman filter estimate. For example, the covariance matrix P provides confidence limits for the estimate $\hat{x}$.

Innovation Property. The one-step ahead prediction error ν is known as the innovation since it represents new information brought into the filter estimate by the measurement y. The innovation has the interesting property that for an optimal filter it is a zero mean Gaussian white noise with covariance matrix Σ. This property can be used to test the optimality of Kalman Filters in applications, to detect changes in the model, to throw out bad data, and to build adaptive and robust Kalman filters. It can be shown that the innovation contains as much information concerning the state as the original measurement sequence.

Numerical Properties. The Kalman filter equations can be written in different forms with different numerical properties. It turns out that equations (2.2.4) - (2.2.8), known as the covariance form of the optimal

filter, though physically easiest to comprehend, are not best suited for numerical computation. In systems with widely separated eigenvalues of the estimation error covariance matrix, round-off errors can lead to non-negative definite covariance matrices. A solution to this problem is obtained by using Cholesky square-roots of the covariance matrices. Consider the $n \times n$ matrix S defined by

$$P = S\,S^T \tag{2.2.9}$$

which is usually computed in either upper or lower triangular form. Equations may be developed for computing S and the covariance matrix is then computed from (2.2.9). Since the condition number of S is half that of P, improved accuracy is obtained on finite word machines. In particular, the accuracy of single-precision square-root Kalman filters is comparable to that of double precision Covariance Kalman filters. For any real S, SS^T is guaranteed to have no negative eigenvalues.

2.2.5 Kalman Filter Approximations in Nonlinear Problems

In reality, the brightness temperature measurements will not be linear in the state. In this case, the Kalman filter equations of Section 2.2.3 do not strictly apply and some approximations must be made. Several techniques have been developed to handle such cases. The Extended Kalman filter uses a linearization approach around the present estimated value of the state. The extended Kalman filter equations for a nonlinear measurement are similar to Eqs. (2.2.3) - (2.2.8) with the following exceptions:

(i) The innovation ν is of the form

$$\nu = y - h(x_0) \tag{2.2.10}$$

instead of (2.2.5).

(ii) The measurement matrix H is replaced by

$$\left.\frac{\partial h(x)}{\partial x}\right|_{x=x_o} \tag{2.2.11}$$

Considerable experience has been gained over the past fifteen years in applying the extended Kalman filter to a variety of problems, most notably in the aerospace field. Results indicate that filter performance is quite good for problems in which $\frac{\partial^2 h}{\partial x^2} P_o < V$. This symbolic inequality means that the expected effect of the second-order terms in the Taylor expansion around the estimate is "smaller" than that of the sensor uncertainty. Thus, good performance may be expected when either the nonlinearity is small or the error variances are small. Although higher-order expansions can be used to form other approximate nonlinear filters, these have generally not worked out in practice due to problems of filter divergence. For more highly nonlinear problems in which the extended Kalman filter does not work well, the best approaches appear to be to either transform the state to make the measurements more linear or use the Iterated Extended Kalman filter. In this approach, the nonlinear measurement function $h(\cdot)$ is re-evaluated at the estimate $\hat{x}$ after processing y. A new estimate is then found and again used to re-evaluate $h(\cdot)$. This process is continued until convergence is achieved.

2.3 Estimation of Discrete-Valued Parameters

Parameters such as cloud type and surface type are conveniently expressed as Boolean variables. That is, they may take only a few discrete (assigned) values. Cloud types may be expressed as "clear" (no clouds), "water cloud" or "ice cloud" and the surface may be categorized as ice, snow or loam, for example. The problem of detecting this type of random variable is distinctly different than estimating continuous random variables, due to the fact that they are discrete-valued.

In the context of the present study, this problem is a simple type of scene classification problem. Its simplicity derives from the fact that the scene is assumed to be homogeneous. Thus, the cloud type or type of surface cover is assumed to be invariant over the entire scene. This assumption is a reasonable one for a study of this type, since feasibility of classification must first be established. Once this has been shown, attention can be given to the more difficult problem of nonhomogeneous scenes. With this assumption, the problem is reduced to one of pattern recognition, which is now discussed briefly.

2.3.1 Classification Using Pattern Recognition

The pattern recognition problem is generally solved in two steps: (1) feature extraction, (2) classification in feature space. Feature extraction may be viewed as a process of selecting a suitably-transformed subset of the raw data for use in the classifier. Ideally, one wishes to keep only the most informative data and throw away the rest. Once the

features are extracted, classification takes place in the selected feature space.

Pattern classification can be accomplished using both parametric and non-parametric methods. For the problems under consideration here, there is expected to be significant overlap between classes and the classes may be non-convex in feature space. This makes it difficult to accurately parametrize the class probability density functions for the purpose of classification. Thus, the use of nonparametric methods is indicated.

There are several significant problems associated with pattern recognition that are of importance in this research. The most significant ones are:

1) finite data base
2) unknown class conditional densities
3) overlapping classes
4) undefined feature space

Finite Data Base

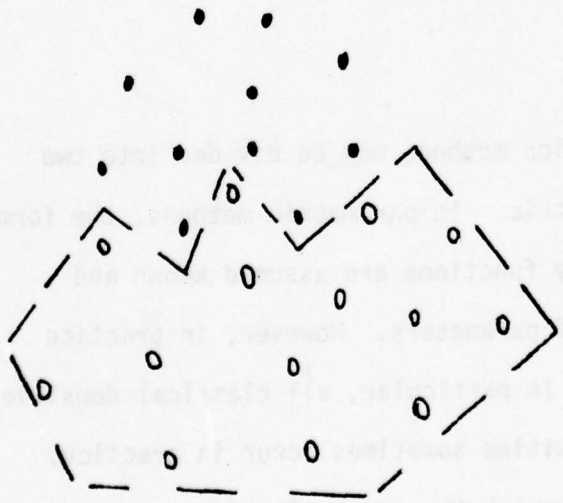
The effect of a finite data base is very significant since the pattern recognition problem is a statistical one. It is important that there is enough data available for training and testing the pattern recognition algorithm to ensure a high degree of confidence in the numerical results obtained from the available data. The exact number of samples required per class is problem-dependent and also depends on the way in which the data is acquired.

Unknown Densities

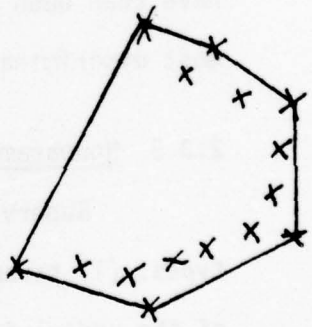
Statistical analysis requires either a set of parameterized probability densities or an accurate set of sample statistics. In the problems to be addressed here, parameterized models are often of questionable validity and nonparametric methods are more appropriate. A variety of methods is available for representing the class-conditional density functions nonparametrically in terms of the actual samples. We have selected a method based on cumulative marginal distribution functions, which are particularly simple to determine.

Overlapping Classes

Another significant problem encountered in pattern recognition is the question of how to handle overlapping classes. In Fig. 2.3.1, two overlapping classes are depicted on the left. The classes are overlapping



overlapping classes



convex hull of
nonconvex class

Figure 2.3.1 Overlapping Classes

since their convex hulls intersect. The decision rule for separating them is not linear, but must be either nonlinear or piecewise-linear.

Undefined Feature Space

There does not appear to be a systematic approach to feature extraction for this problem. One approach is to use intuitively-derived features and these have much merit since they are based on experience and the human pattern recognition capability. However, the digital computer is capable of performing some operations of nonlinear transformation, correlation and manipulation of large amounts of data which cannot be matched by the human. This computing power should be brought to bear in multivariate multiclass problems in order to perform data compression and feature extraction for optimal and accurate classification. The approach used here has utilized a large set of potential features which have then been analyzed statistically to determine the ones with the most discriminating power.

2.3.2 Nonparametric Decision Rules

Supervised pattern recognition methods may be divided into two types: (1) parametric, (2) nonparametric. In parametric methods, the form of the underlying probability density functions are assumed known and are parametrized by a small number of parameters. However, in practice this assumption is usually suspect. In particular, all classical densities are unimodal, whereas multimodal densities sometimes occur in practice. The use of nonparametric methods, in which the power of modern computers is more fully utilized, is suggested.

The nonparametric classification problem may be stated as follows. An n-dimensional random vector x of observations, or features, is assumed to belong to one of m classes $\Gamma_1, \Gamma_2, \ldots, \Gamma_m$ characterized by a set of unknown class-conditional probability density functions $p(x|\Gamma_1), \ldots, p(x|\Gamma_m)$, where $p(A|B)$ denotes the probability of A given that B has occurred. The m classes have known prior probabilities $\pi_1, \pi_2, \ldots, \pi_m$. We are given a set of N labeled vectors $x = \{x_1, x_2, \ldots, x_N\}$. The problem is to determine a decision rule on the basis of x and the labels for discriminating between the m classes.

There are several different approaches to nonparametric pattern recognition. One approach is to estimate the class conditional density functions $p(x|\Gamma_i)$ from the data using the methods of Parzen (1962), Wagner (1975) or Loftsgaarden and Quesenberry (1965), for example. Another procedure is to estimate the a posteriori class probabilities $p(\Gamma_i|x)$ directly from the data. This approach is closely related to the k-nearest neighbor decision rules which do not compute $p(x|\Gamma_i)$ but produce a decision rule directly. This latter approach is closest in spirit to the method used in this study.

2.3.2.1 k-Nearest Neighbor Rules

Perhaps the most popular nonparametric decision rules are the k-nearest neighbor rules first investigated by Fix and Hodges (1951). Training samples from the m populations are gathered together into a single population, with each member labeled according to the class from which it originated. Now suppose it is desired to classify a new point, x, on the basis of the data in the training set. The k closest points to x are found and x is assigned to the class with the largest representation in the set of k points. Fix and Hodges were able to show that this rule is asymptotically Bayes efficient as k and N, the number of samples in the training

set, approach infinity if

$$\lim_{N \to \infty} k(N) = \infty$$

$$\lim_{N \to \infty} \frac{k(N)}{N} = 0$$

Cover and Hart (1961) have investigated this rule for fixed k and derived bounds for the simplest case of k = 1 (one-nearest neighbor decision rule). The result is that the error rate P_e for m populations is bounded asymptotically by

$$P_e^* < P_e < P_e^* \left(2 - \frac{mP_e^*}{m-1}\right)$$

where P_e^* is the Bayes error rate. This result implies that $P_e < 2P_e^*$, independent of the number of populations.

2.3.2.2 Nonparametric Partitioning

Partitioning using hierarchical decision trees is another approach to nonparametric supervised learning. This approach has been employed by several workers. Henrichon and Fu (1969) considered the problem of partitioning the real line on the basis of labeled samples. They allowed partitions associated with each class to be disjoint. That is, a single class was not necessarily constrained to a single interval. The multidimensional problem was handled by considering each dimension independently and then building up a total score on the basis of classification in each dimension. Meisel and Michalopoulos (1973) attacked the problem of finding decision trees which tend to minimize the average number of comparisons to arrive at a given decision. They developed a dynamic programming solution to the problem. Unfortunately, the number of trees which must be evaluated

is of the order of $N^3/6$ where N is the total number of samples. Thus, for even moderate-sized data bases, dynamic programming solutions are not possible. For example, if N = 1000, then $10^9/6$ trees would have to be evaluated.

2.3.2.3 A New Partitioning Rule

The method used in this study is a new nonparametric partitioning decision tree rule developed by Friedman (1976). This method has been applied successfully to electrocardiogram classification by Gustafson et al. (1978) and appears to have strong general merit. It has the following desirable properties:

(i) The procedure is computationally simple for both training and classification phases. This is of great importance in the present problem, due to its high dimensionality.

(ii) The decision rule that results from the training phase is asymptotically Bayes' risk efficient as the number of training samples increases without limit.

(iii) Additional features may be easily added prior to or during the training phase. These may be linear or nonlinear. Non-informative features are simply ignored in the training phase. However, any of the additional features which are informative are used, and become part of the partitioning algorithm.

The severest limitation is that it was derived for the two-class problem only. Although it was suggested that the M-class problem be treated as a series of M two-class problems, this is not completely straightforward in practice. Nevertheless, the method was selected as the best overall approach to the problem.

For the purposes of presenting the two class problem, we will designate the two classes 1 and 2. Further, we will use the fact that $\ell_1\pi_1 = \ell_2\pi_2$ where ℓ_i is the cost of misclassifying class i and π_i is the a priori probability of class i. Extensions to a more general set of cost functions are straightforward and will be presented later in this section. We will also assume that associated with both classes is a set of features $\{x_i\}$. In the problem of multispectral classification, these features may be radiances, brightness temperatures, or their gradients, ratios, second differences, etc. That is, the features may consist of any function of the observed data that we suspect of significance in discriminating the classes.

Given a single feature, x_i, of this set, we now seek a simple decision rule that partitions the classes on the value of x_i. That is, we desire the value of x_i, say x_i^*, such that we place the observation into one class if $x_i < x_i^*$ and the other if $x_i > x_i^*$ in such a way as to minimize the probability of error. For a given feature, x_i, it may be shown that the value x_i^* is the value that maximizes the distance between the marginal cumulative distribution functions (CDF) or Kolmogorov-Smirnov Distance (KS Distance). That is, if $F_1(x_i)$ is the marginal CDF of Class 1 in x_i space and $F_2(x_i)$ the CDF of Class 2 in x_i space, x_i^* maximizes the quantity:

$$D_i(x_i) = |F_1(x_i) - F_2(x_i)| \qquad (2.3.1)$$

where $D_i(x_i)$ is the KS Distance. To show that this is the appropriate decision point, let us assume that $F_1(x_i) > F_2(x_i)$ for all x_i and that the decision rule is based on some point x_i'. This rule will be to choose Class 1 if $x_i < x_i'$ and choose Class 2 if $x_i > x_i'$. Then the expected cost of a decision is:

$$C_i = \ell_1\pi_1 \Pr\{x_i > x_i' | \text{Class 1}\} + \ell_2\pi_2 \Pr\{x_i' | \text{Class 2}\} \qquad (2.3.2)$$

Since it is assumed that $\ell_1\pi_1 = \ell_2\pi_2 = k$ this cost is

$$C_i = k[(1 - F_1(x_i')) + F_2(x_i')]$$

$$= k[1 - D_i(x_i')] \qquad (2.3.3)$$

The cost is thus minimal at the x_i' that maximizes $D_i(x_i)$.

If a number of features x_i exist, the optimum choice of a feature on which to partition the data is the one that yields the minimum cost C_i. By the preceding analysis, this feature is the one with the maximum $D_i(x_i^*)$.

Having partitioned the original set of observations into two new sets on the basis of one feature, it is now possible to reapply this partitioning rule to each of the new sets if further discrimination is warranted.

The decision as to what constitutes a terminal node is somewhat arbitrary. Since the actual CDF's of the features will be unknown in general, the algorithm must rely on sample CDF's to perform the partitions. The confidence intervals attached to such sample statistics are a function of the number of observations entering into their computation. The definition of a terminal node has been historically taken to be a node that contains only members of a single class or a node that has too few observations to allow the CDF's to be meaningful.

The label on a terminal node or leaf is given by the minimum cost decision at that node. For each class, there is a probability that a sample will arrive at the leaf j. The cost of labeling a leaf class 1 is thus:

$$C(\text{decide 1 at node j}) = \ell_2\pi_2 \frac{\Pr(\text{arrive at leaf j} | \text{class 2})}{\Pr(\text{arrive at leaf j})}$$

and the cost of labeling a leaf class 2 is

$$C(\text{decide 2 at node j}) = \ell_1\pi_1 \frac{\Pr(\text{arrive at leaf j} | \text{class 1})}{\Pr(\text{arrive at leaf j})}$$

The node is labeled with the class that gives the lower cost.

For the purposes of the study undertaken in this section, the terminal nodes were defined by one of two criteria. A node was considered terminal if either:

i) only one class appears at a node or

ii) the maximum distance $D_i(x_i^*)$ does not reject the null hypothesis of the Kolmogorov-Smirnov two sided - two sample test at greater than the 90 percent confidence level.

The first criterion is an obvious one. The second, however, warrants some further discussion.

The Kolmogorov-Smirnov two sided - two sample test is a powerful yet simple nonparametric test for hypothesis that two samples are drawn from different populations (e.g., Conover, 1971). The test is based on order statistics and uses the maximum KS Distance $D_i(x_i^*)$ as its sufficient statistic. Since this statistic is an integral part of the tree building algorithm, the test is easily implemented. The null hypothesis of the test is that the two samples are drawn from the same population. This hypothesis is rejected if the $D_i(x_i^*)$ is greater than some value that depends on the sample sizes and the required confidence. The rationale of using this test as a test for a terminal node is that if the algorithm is going to partition the classes with regard to some feature, the difference between the populations along that feature should be real to some ascribable confidence.

One important quality of the Friedman Tree algorithm is its invariance under monotone transformations of the features. That is, the features used to partition the classes at each node in the tree will remain the same no

matter what monotone transformation is applied to them. In cases such as the one at hand in which the features are positive numbers, this means that the tree produced will be the same for the choice of a basic feature x_i as it will for the choice of x_i^2, x_i^3, $\log(x_i)$, $1/x_i$, $-x_i$, $1/(1+x_i)$, e^{x_i}, etc. as basic features. The only difference between the various trees will be the value of the cut at each node.

For the more general case of $\pi_1 \ell_1 \neq \pi_2 \ell_2$, the algorithm must be modified slightly. Since equation 2.3.2 does not reduce to equation 2.3.3, equation 2.3.2 must be used directly. That is, for a given feature i, the choice of a point on which to partition is the one that yields the minimum cost:

$$C_i = \ell_1 \pi_1 \, (1 - F_1(x_i')) + \ell_2 \pi_2 \, F_2(x_i') \qquad (2.3.4)$$

and the optimum feature is the one with the minimum C_i. The balance of the algorithm including the terminal node definition remains as before.

For the M-class discrimination problem, it is possible to use the Friedman Tree approach by considering this to be M two class problems. That is, for each of the M classes a tree is built with a specific class being class 1 and the remaining classes being class 2. Each leaf of the various trees are labeled with the cost of deciding that the measurement was a member of the class 1 of the respective tree. A measurement is then identified with the class whose tree has the lowest class 1 leaf cost.

2.4 Development of Inversion Algorithm for Continuous-Valued Parameters

In this section, the methods of Section 2.2 are specialized to the problem of determining continuous-valued parameters under certain important assumptions needed to solve for the meteorological parameters of interest. We have considered the following inversion methods:

(1) linear regression

(2) nonlinear regression

(3) extended Kalman Filter

(4) iterated extended Kalman Filter

(5) modal estimator

The first two methods, based on regression models, are distinctly different in philosophy than the last three methods, which are based on Bayesian statistical inference. These two philosophies are now discussed and compared.

2.4.1 Regression Methods

The use of regression analysis is a powerful tool in analyzing the relationship between sets of variables of different types. In the present problem, linear and nonlinear regression models for solving the inversion problem have been developed.

In computing regression models, it is important to determine the effect of using different subsets of independent variables. It is desirable to use only those variables which significantly reduce the regression errors. Overfitting can lead to models which are overly sensitive to noise. Stepwise regression provides a partial automation of the variable selection

process. It is based on a technique which in the process of computing an ordinary regression on m predictors obtains, at essentially no extra expense, m intermediate regressions which are useful in determining functional relationships between the dependent variable and several selected subsets of the total set of independent variables, or predictors. In the simplest case, one variable is added at each step. There are several meaningful statistical criteria which may be used to determine which predictor variable to add to the model:

(i) add the predictor whose F-to-enter statistic has the largest value;

(ii) add the predictor that gives the greatest decrease in the residual sum of squares;

(iii) add the predictor which gives the greatest increase in the multiple correlation between the dependent variable and the predictors.

The F-to-enter statistic for a predictor is the F-statistic for testing the significance of the regression coefficient the predictor would have if it were added. All three of these criteria are, in fact mathematically equivalent; however the F-to-enter statistic was used in programs utilized in this study.

In addition to adding variables it is also desirable to provide for removal of variables. It may happen that recently added variables may in combination render a variable entered at an earlier state statistically insignificant. Removal of variables is based on an F-to-remove test.

In the stepwise regression procedure, the test for removal of predictors is first made for each predictor already in the model. If a predictor is removed, the program proceeds to the next step. If no predictor is removed, the step continues with the addition of a predictor. The stepwise regression procedure terminates when no predictor is either deleted or added at a step. The details of the stepwise regression procedure may be found, for example, in Draper and Smith (1966).

The regression method used throughout this study was performed by the program RLSEP, which was available at AFGL as part of the International Mathematical Statistics Library (IMSL).

It should be pointed out that the statistical D method which has been used in several meteorological parameter inversion studies, is a regression technique.

2.4.2 Bayesian Estimation

We recall the Bayes' formula

$$p(x|y) = \frac{p(y|x)\ p(x)}{\int_{\Omega_x} p(y|x)\ p(x)\ dx} \qquad (2.4.1)$$

In order to compute the conditional density $p(x|y)$ we need:

(a) the priori density $p(x)$

(b) the conditional density $p(y|x)$, which can be computed from the model $y = h(x) + v$ if the density of v is known.

Given $p(x|y)$ we have considered two different types of estimators: (1) minimum variance, (2) maximum likelihood.

Minimum-Variance Estimate

The minimum-variance estimate $\hat{x}_{mv}$ is the conditional mean

$$\hat{x}_{mv} = \int_{\Omega_x} x\, p(x|y)\, dx \qquad (2.4.2)$$

and can be computed, for general nonlinear problems, only by multiple integration over the domain Ω_x. This is usually not computationally feasible for more than two variables. In the present problem, it is not feasible to evaluate (2.4.2) directly and some approximations must be made. The most generally useful method is based on the iterated extended Kalman Filter discussed in Section 2.2.5.

Maximum Likelihood Estimate

The most likely value of x is given by the mode

$$\hat{x}_{ml} = \arg \max_{x \in \Omega_x} p(x|y) \qquad (2.4.3)$$

and is sometimes referred to as the maximum likelihood Bayes estimate. $\hat{x}_{ml}$ can, in principle, be found more easily than $\hat{x}_{mv}$, since integration over Ω_x is not required. However, for nonlinear problems, exhaustive search techniques may have to be used to find the global maximum of $p(x|y)$.

2.4.3 Comparison of Bayesian and Regression Methods

It is instructive to compare Bayesian and regression methods to assess the fundamental differences. This is done in Tables 2.4.1 and 2.4.2. Note that both methods require a model of the system. The Bayesian approach uses a forward model which is based on the physics of the problem (i.e., a model to predict intensity or brightness temperature is utilized). Additive

TABLE 2.4.1

Summary of Regression and Bayesian Techniques

for Inversion

Measurement : $y = h(x) + v$; p_v = prob. density for v.

Regression	Bayes
$\hat{x} = g_r(y)$	$\hat{x} = g_B(y;h,p_v)$
g_r determined by data fitting to selected data base. Data must be statistically representative.	g_B determined to minimize selected performance measure. May depend on form of h and p_v

TABLE 2.4.2

COMPARISON OF BAYESIAN AND REGRESSION METHODS

REGRESSION	ISSUE	BAYES
NOT REQUIRED	FORWARD MODEL	REQUIRED. USES PHYSICS AND/OR REGRESSION ANALYSIS
REQUIRED. BASED ON STATISTICS ONLY	INVERSE MODEL	NOT REQUIRED. USES FORWARD MODEL AND STATISTICS OF NOISE
DIFFICULT TO PERFORM	DATA REASONABLENESS MODEL	EASILY PERFORMED USING MEASUREMENT RESIDUALS
ESTIMATED FROM A PRIORI STATISTICS	CONFIDENCE IN ESTIMATES	COMPUTED ON-LINE, ADAPTIVELY
DIFFICULT TO HANDLE	MISSING DATA	EASY TO HANDLE

noise is included to account for two effects:

(1) sensor noise;

(2) modeling errors due to uncertainty in the function $h(\cdot)$.

Once the forward model is known, estimates are found as a function of the measurement y, the mean and covariance of the prior estimate (m, P_o) and the covariance of measurement noise (V).

In contrast, regression methods use an inverse model, which is not physically-based, but rather must be generated statistically from data. Noise may be included in the model, as with the statistical-D method. The Bayesian technique has advantages in handling missing data and checking for bad data. Since the covariance of the measurement residuals is known, (c.f. Eq. (2.2.7)), it is simple to derive statistical outlier tests to throw out bad data.

Another significant advantage of the Bayesian approach would accrue during on-line data processing in which parameters are estimated, or tracked, over time. The Bayesian estimators we propose to use not only provide estimates of the desired parameters but also provide estimates of the covariances of estimation errors, since the full estimation error covariance matrix is computed as part of the Kalman Filtering process. This matrix is used to optimally weight the information coming in via the measurements relative to the information contained in all past measurements.

2.4.4 Development of Bayesian Inversion Methods

2.4.4.1 Extended Kalman Filter

The state, x, of interest contains all of the parameters we wish to estimate. It is assumed to have mean m and covariance matrix P_o. The measurement is of the form

$$y = h(x) + v \tag{2.4.4}$$

with v a zero-mean random variable with covariance matrix V, independent of x.

The form of the estimator is

$$\hat{x} = m + M\,[y - \hat{y}] \tag{2.4.5}$$

where $\hat{y} = h(m)$ is the predicted value of y and $\nu = y - \hat{y}$ is the measurement residual, the difference between the actual and expected measurement. If $y = h(m)$ there is no information contained in y relative to m and $\hat{x} = m$. The matrix M, referred to as the Kalman gain matrix, is to be found to yield the best performance.

Defining the estimation error

$$e = \hat{x} - x \tag{2.4.6}$$

we have the error covariance matrix

$$P = E[e\,e^T] \tag{2.4.7}$$

where $E[\cdot]$ is the ensemble average of $[\cdot]$ and $(\cdot)^T$ is the transpose of $(\cdot)$.

The objective of a minimum-variance estimator is to choose M such that $u^T P u$ is minimized for any real non-zero vector u of the same dimension as x. If $h(\cdot)$ is linear, the solution is the classical Kalman Filter (1960), given in Eqs. (2.2.4) - (2.2.8). For nonlinear $h(\cdot)$, we must resort to approximate solutions.

The extended Kalman Filter is one approach to nonlinear problems and is based on a linearized measurement equation. Thus we assume

$$h(x) = h(m) + \left.\frac{\partial h(x)}{\partial x}\right|_{x=m} (x-m) \tag{2.4.8}$$

which gives

$$y = h(m) + H(m)(x-m) + v \tag{2.4.9}$$

where

$$H(m) = \left.\frac{\partial h(x)}{\partial x}\right|_{x=m} \tag{2.4.10}$$

If we define a new measurement

$$\tilde{y} = y - h(m) - H(m)\, m \tag{2.4.11}$$

then

$$\tilde{y} = H(m)\, x + v \tag{2.4.12}$$

$$\hat{\tilde{y}} = H(m)\, m \tag{2.4.13}$$

and the measurement y is linear in x. Applying the Kalman Filter equations [cf. (2.2.4) - 2.2.8)] to this problem gives the following solution

$$\hat{x} = m + M^* \nu \tag{2.4.14}$$

$$\nu = y - h(m) \tag{2.4.15}$$

$$M^* = P_o\, H(m)^T\, \Sigma^{-1} \tag{2.4.16}$$

$$\Sigma = H(m)\, P_o\, H(m)^T + V \tag{2.4.17}$$

$$P = P_o - M^*\, H(m)\, P_o \tag{2.4.18}$$

where M^* is the optimal gain matrix, Σ is the estimated covariance matrix for the residuals ν and P is the estimated covariance matrix for the a posteriori estimation errors.

This estimator will perform reasonably well in problems which are not highly nonlinear. The effect of the nonlinearity can be explicitly demonstrated by computing $h(\cdot)$ to second order in (2.4.8). Let $e = \hat{x} - x$, $e_o = x - m$ and assume y is a scalar for simplicity. Then, to second order

$$e = e_o + M\left[-H(m)\, e_o + \frac{1}{2} e_o^T S(m)\, e_o + v\right] \tag{2.4.19}$$

where

$$[S(m)]_{ij} = \frac{\partial^2 h(x)}{\partial x_i \partial x_j} \tag{2.4.20}$$

Then, defining P^+ to be the actual covariance matrix of errors

$$\begin{aligned} P^+ &= E[e\, e^T] \\ &= [I - M\, H(m)]\, P_o\, [I - M\, H(m)]^T \\ &= M\left[V + \frac{1}{4} E\{(e_o^T S(m)\, e_o)^2\}\right] M^T \end{aligned} \tag{2.4.21}$$

It can be shown that if $S(m) = 0$ and $M = M^*$, then $P^+ = P$. Equation (2.4.21), good to second order for any M, shows explicitly that $E\{e_o^T S(m)\, e_o)^2\}$ must be small compared to V for the extended Kalman Filter to be a good estimator. If we write

$$\tilde{\Sigma} = \Sigma + E\{(e_o^T S(m)\, e_o)^2\} \tag{2.4.22}$$

and use $\tilde{\Sigma}$ in place of Σ in (2.4.17), then M^* and P would be better approximations and the filter would be expected to perform with more accuracy.

An approach which has generally been found to yield better results in practice was developed by Denham and Pines (1966) and is called the Iterated Extended Kalman Filter.

2.4.4.2 Iterated Extended Kalman Filter

In equation (2.4.21), it was shown how nonlinearity can affect the Extended Kalman Filter. A major reduction in estimation error can be made by a simple modification of the Extended Kalman Filter. The modification consists of writing $\hat{y}$ in (2.4.5) as

$$\hat{y} = h(m)$$

$$= h(\hat{x}) + H(\hat{x})\,(m - \hat{x}) \tag{2.4.23}$$

for any estimate $\hat{x}$. For nonlinear $h(\cdot)$ it is expected that $y - \hat{y}$ will be smaller using (2.4.23) for $\hat{y}$ rather than using $\hat{y} = h(m)$, especially if the nonlinearities are significant. This observation provides the motivation for the following iteration:

(i) Set $\beta^{(0)} = m$

(ii) Compute

$$\beta^{(k+1)} = m + M^{(k)}\,\nu^{(k)} \tag{2.4.24}$$

$$\nu^{(k)} = y - h(\beta^{(k)}) - H(\beta^{(k)})\,(m - \beta^{(k)}) \tag{2.4.25}$$

$$H(\beta^{(k)}) = \frac{\partial h(x)}{\partial x}\bigg|_{x=\beta^{(k)}} \tag{2.4.26}$$

$$M^{(k)} = P_o\,H(\beta^{(k)})^T\,[\Sigma^{(k)}]^{-1} \tag{2.4.27}$$

$$\Sigma^{(k)} = H(\beta^{(k)}\,P_o\,H(\beta^{(k)})^T + V \tag{2.4.28}$$

(iii) Compute

$$\eta^{(k)} = \|\beta^{(k+1)} - \beta^{(k)}\|$$

for some norm $\|\cdot\|$

(iv) If $\eta^{(k)} > \varepsilon$, return to (ii); otherwise continue

(v) Compute new state estimate as

$$\hat{x} = \beta^{(k+1)} \tag{2.4.29}$$

(vi) Update estimation error covariance matrix

$$P = P_o - M^{(k+1)} H(\beta^{(k+1)}) P_o \tag{2.4.30}$$

The effect of using this scheme can be evaluated by comparing the estimation error covariances for the iterated filter to that of the non-iterated filter (eq. (2.4.21)). Again we assume a scalar measurement. Then, to second order

$$\begin{aligned} e^{(k+1)} &= x - \beta^{(k+1)} \\ &= e_o + M^{(k)} \left[-H(\beta^{(k)}) e_o + \frac{1}{2}(e^{(k)})^T S(\beta^{(k)}) e^{(k)} + v\right] \end{aligned} \tag{2.4.31}$$

Comparing this with (2.4.19), which is the Extended Kalman Filter results shows that a significant reduction in the nonlinear term has been made, due to the fact that $e^{(k)}$ will almost always be smaller than e_o if the algorithm converges.

2.4.4.3 Estimation of the Mode

From (2.4.1) and (2.4.3), the mode of the a posteriori density $p(x|y)$ is given by

$$x_m = \arg \max_{x \in \Omega_x} p(y|x)\, p(x) \tag{2.4.32}$$

and is equivalent to the Bayes maximum likelihood estimate. In order to find $\hat{x}_m$, we need to specify the probability densities of x and the noise v. We assume x to be normally distributed with mean m and covariance matrix P_o and assume v to be normally distributed with zero mean and covariance

matrix V. Thus, if x is of dimension n and y is of dimension q:

$$p(x) = \frac{1}{(2\pi)^{n/2}|P_o|} \exp\left\{ -\frac{1}{2} (x-m)^T P_o^{-1} (x-m) \right\} \tag{2.4.33}$$

$$p(y|x) = \frac{1}{(2\pi)^{q/2}|V|} \exp\left\{ -\frac{1}{2} (y-h(x))^T V^{-1} (y-h(x)) \right\} \tag{2.4.34}$$

where $|(\cdot)|$ denotes the determinant of $(\cdot)$. Neglecting constant terms and noting that the logarithm is a monotone function, we see that (2.4.32) is equivalent to

$$x_m = \arg \max_{x \in \Omega_x} \phi(x) \tag{2.4.35}$$

where

$$\phi(x) = (x-m)^T P_o^{-1} (x-m) + (y-h(x))^T V^{-1} (y-h(x)) \tag{2.4.36}$$

At the mode

$$\frac{\partial \phi(x)}{\partial x}\bigg|_{x=\hat{x}_m} = 0 \tag{2.4.37}$$

or

$$\hat{x}_m = m + P_o H(\hat{x}_m)^T V^{-1} [y - h(\hat{x}_m)] \tag{2.4.38}$$

where

$$H(\hat{x}_m) = \frac{\partial h(x)}{\partial x}\bigg|_{x=\hat{x}_m} \tag{2.4.39}$$

Equation (2.4.38) can be solved, in principle, in a variety of ways. However, we note here that the appropriate method to be used may depend upon the functional form of $h(\cdot)$ as well as the value of the initial guess.

III. INVERSIONS USING MICROWAVE DATA FOR CONTINUOUS-VALUED PARAMETERS

3.1 Introduction

It is convenient to assemble the parameters we wish to estimate together into a single state vector x:

$$x = \begin{bmatrix} \text{cloud parameters} \\ \text{surface parameters} \\ \text{atmospheric parameters} \end{bmatrix}$$

We are given an *a priori* estimate of x in the form

$$\hat{x}_0 = f(\text{latitude, season, forecasts, etc.})$$
$$= f(\alpha)$$

Here a caret (^) denotes an estimate, $f(\cdot)$ is a known function obtained from *a priori* data analysis and α is used to represent a vector of parameters used to predict x prior to incorporating the microwave data.

The next step is to process the microwave data to find a better estimate of x. Assume that a total of m microwave channels are used and that all data is obtained simultaneously. If the measurements are collected into a m-dimensional vector y then the problem is to find a "best" estimator for x of the form

$$\hat{x} = g(\hat{x}_0, y)$$

Before presenting the problem solutions which have been investigated, we will give a brief description of the method of approach.

3.2 Method of Approach

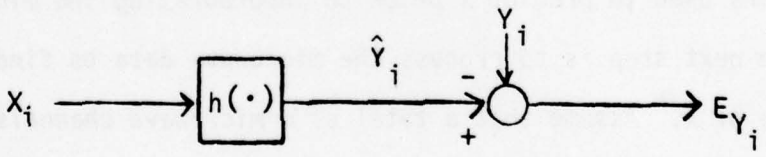
The method of approach which has been used for inversion of microwave data may be summarized as follows:

(A) GENERATE DATA BASE USING UNCONDITIONAL STATISTICS OF THE STATE

(B) FORWARD SIMULATION

$\{X_1, X_2, \ldots\}$ SELECTED TO ACCOUNT FOR EXPECTED STATISTICAL VARIATIONS IN X

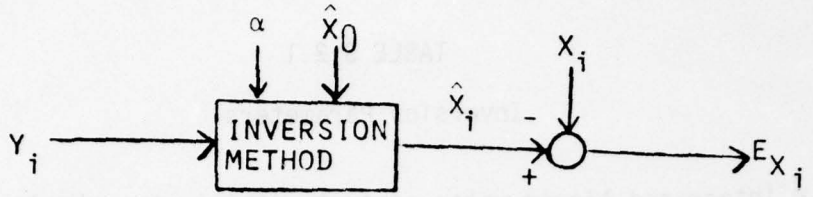
(C) COMPUTE FORWARD MODEL



FIND $h(\cdot)$ TO MINIMIZE SELECTED NORM ON $\{E_{Y_1}, E_{Y_2}, \ldots\}$;

GENERALLY, USE LEAST SQUARES.

(D) DEVELOP INVERSION TECHNIQUES



E_{X_i} = INVERSION ERROR

(E) EVALUATION

- ANALYZE MODELING ERRORS $\{E_{Y_i}\}$
- ANALYZE INVERSION ERRORS $\{E_{X_i}\}$
- DESIGN OF TRAINING/TEST SETS
- EFFECT OF MISSING DATA

3.2.1 Data Base

The data bases were constructed so as to cover the expected variations in the independent variables. The independent data were determined on the basis of discussions with ERT and AFGL personnel. Three different data bases were used, which are described in the sequel. The basic parameters of interest are given in Table 3.2.1.

3.2.2 Forward Model Computation

Each data base was used to generate a set of forward models for predicting microwave brightness temperatures, in °K, from the independent variables. Four different channels were run: 10.7 GHz, 19 GHz, 37 GHz,

TABLE 3.2.1

Inversion Parameters

LWC = integrated liquid water content within water clouds
(milligram/cm^2)

WS = wind speed over the ocean surface
(m/sec)

$$WST = \begin{cases} 8 & ; WS < 8 \\ WS & ; 8 < WS < 22.5 \\ 22.5 & ; WS > 22.5 \end{cases}$$

DLT = atmospheric temperature deviation (°K), assumed constant at all altitudes

CT = cloud thickness
(hundreds of meters)

CH = cloud top height
(hundreds of meters)

WV = integrated atmospheric water vapor
(gm/cm^2)

R = rainfall rate
(mm/hr)

and 94 GHz using both linear and nonlinear regression. The best models are generally quite nonlinear. The F ratios, goodness-of-fit (R^2) and rms regression error (σ) were evaluated to assess the characteristics of the regression models.

3.2.3 Inversion Techniques

The different types of inversion techniques investigated during the study have been discussed in Section II. There are two basic types of techniques: (1) regression, (2) Bayesian. These were compared in Table 2.4.2 and it was shown that Bayesian techniques offer more flexibility in data handling and modeling than regression techniques do.

The emphasis in the present work is on Bayesian analysis. The basic ideas employed in the Bayesian approach have been outlined in Section 2.4. The basic approach has been utilized to determine several different types of estimates. Minimum variance estimates are desirable to achieve, but in practice are very difficult to compute for nonlinear problems. No general methods presently exist for computation of the conditional mean in nonlinear problems. For this reason, it is common in practice to rely on other types of estimates. The maximum likelihood estimate can be found in principle by solving a nonlinear programming problem; i.e., one can actually compute maximum likelihood estimates. However, the disadvantage of these estimates is that no prior information is used. This is an important consideration in the present problem, since a priori estimates carry a significant amount of information relative to the desired estimates. The mode of the a posteriori density is an attractive estimate, since it contains

the a priori information and can be computed in practice, subject to certain moderate restrictions. The iterated-extended Kalman filter described in the sequel can be viewed as an approximate modal estimator for the a posteriori density.

3.3 Inversions for Data Set #1

3.3.1 Data Base

The first data set used for numerical studies is summarized in Table 3.3.1 and consisted of 803 simulations using the GABTAWF program. The data in the table were used to evaluate the performance of the inversion techniques. In order to achieve better regression model fits in the presence of errors, a total of 60 additional cases were utilized in the fitting set which consisted of rarely-occurring cases in which the clouds had an unusually high liquid water content. This also lends to reduce the fit error at outlying points in the test set. The additional cases consisted of ten values of LWC (.286, .312,..., .520) with the following six conditions on the other variables:

	WST	DLT	CT	CH	W V
1.	8	0	25	49	1.932
2.	8	0	25	49	1.375
3.	20	0	25	49	1.375
4.	8	0	25	49	2.442
5.	20	0	25	49	2.442
6.	20	0	25	49	1.932

All data were run for Midlatitude Spring conditions over the ocean with the sun overhead. A summary of the estimated parameters and techniques to be evaluated are given in Table 3.3.2.

Wind speed (WST)		Δ Temperature (DLT)		Cloud thickness (CT)		Cloud top height (CH)		Integrated water vapor (WV)	
m/s	#	°k	#	m	#	m	#	gm/cm^2	#
8.	374	-5.	80	0	45	660	40	.819	10
8.5	20	0.	683	330	40	1000	9	.893	40
10.	40	5.	40	500	602	1320	40	.90	10
12.5	20			660	40	1500	100	1.016	40
15.	40			1000	4	2000	9	1.575	5
17.5	20			1400	2	2900	310	1.621	80
20.	269			1500	2	3000	9	1.704	262
22.5	20			2000	4	3400	2	1.788	2
				2500	64	3900	2	1.807	40
						4000	9	1.819	100
						4400	2	1.837	2
						4900	262	1.850	2
						5000	9	1.895	2
								1.919	40
								1.932	42
								2.009	2
								2.225	2
								2.268	10
								2.288	40
								2.336	40
								2.442	20
								2.485	2
								2.5	10

TABLE 3.3.1(a)

Independent Data Used for Numerical Studies - Data Base #1

TABLE 3.3.1(b)

Independent Data Used for Numerical Studies - Data Base #1

Integrated Liquid Water Content	
gm/cm^2	#
.006,(.006),.120*	600
.0043,(.0043),.086	40
.0076,(.0076),.152	40
.026,(.026),.260	60
0	45
.009	2
.0165	4
.0240	4
.0315	4
.0390	4
total	803

* a,(b),c denotes values starting at a with increment b and final value c

TABLE 3.3.2
DATA FOR NUMERICAL EXPERIMENTS

MIDLATITUDE SPRING

SUN OVERHEAD

OVER SEA

STATE:

$$X = \begin{bmatrix} \text{INTEGRATED LIQUID WATER} \\ \text{CLOUD THICKNESS} \\ \text{CLOUD TOP HEIGHT} \\ \text{SURFACE WIND SPEED} \\ \text{ATMOSPHERIC TEMPERATURE} \\ \text{INTEGRATED WATER VAPOR} \end{bmatrix} = \begin{bmatrix} \text{LWC} \\ \text{CT} \\ \text{CH} \\ \text{WST} \\ \text{DLT} \\ \text{WV} \end{bmatrix}$$

WST = TRUNCATED WIND SPEED
(WST > 8 M/SEC)

DLT = CONSTANT TEMPERATURE VARIATION AT ALL HEIGHTS

ERROR = TRUE - ESTIMATED

ESTIMATES:

OL ≃ OPEN LOOP (USE A PRIORI MEAN)

L ≃ LINEAR REGRESSION

NL ≃ NONLINEAR REGRESSION

EKF ≃ EXTENDED KALMAN FILTER

IEKF ≃ ITERATED EXTENDED KALMAN FILTER

3.3.2 Forward Models

The forward regression models for three microwave channels (19, 37, 94 GHz) are shown in Tables 3.3.3 - 3.3.5 and are seen to be highly nonlinear. The 19 GHz temperature predictors of most significance are integrated liquid water and water vapor. Less significant predictors are wind speed (i.e., foam coverage), cloud thickness - height product and the product of cloud height and integrated liquid water content. The most significant predictor for the 37 CHz channel is the product of integrated liquid water content and wind speed. Water vapor is seen to have little effect. At 94 GHz, the square of integrated liquid water content, alone and in combination with cloud thickness, is seen to be a strong predictor. Other combinations of these two variables are seen to be of significance.

3.3.3 Inversion Results

Regression models for direct inversion for the parameters are given in Tables 3.3.6 - 3.3.10 and are seen to be highly nonlinear. In the tables, m_d is the dependent mean and σ_{prior} is the rms variation of the parameter over the 863 cases in the data base.

The inversion errors using two different regression models are summarized in Table 3.3.11. A 16-th order model for LWC yields somewhat better results than an 11-th order model (Table 3.3.6); however, the results for the other parameters are almost identical.

Inversion errors for the extended and iterated extended Kalman Filter are shown in Table 3.3.12. The parameter σ_r is the rms channel noise, in °K, utilized in the filter. Thus the measurement covariance matrix V is either

TABLE 3.3.3

FORWARD MODEL FOR PREDICTING 19 GHz TEMP.

(UNITS GIVEN IN TABLE 3.2.1)

T19

REGRESSION VARIABLES

		COEFF.	F
1	LWC	.16	5479
2	WST	-2.9	1412
3	DLT	-.27	208
4	CT	.08	49
5	CH	-.016	12
6	WV	6.1	5414
7	CT·CT	-12 E-5	985
8	CT·CH	-20 E-4	1516
9	LWC·CT	-21 E-4	427
10	LWC·CH	25 E-4	1452
11	LWC/CT	0	0
12	WST·WST	.165	355
13	CT-CH	0	0
14	WV·WV	0	0
15	LWC·LWC·CT	0	0
16	LWC·LWC·CH	0	0

$R^2 = .9976$

$\sigma = 0.93°c$

INTERCEPT: 135.7

TABLE 3.3.4

FORWARD MODEL FOR PREDICTING 37 GHz TEMP.

(UNITS GIVEN IN TABLE 3.2.1)

T37

REGRESSION VARIABLES

		COEFF.	F
1	LWC	.53	634
2	WST	-2.2	255
3	DLT	-.46	198
4	CT	0	0
5	CH	-.036	9.8
6	WV	3.19	473
7	LWC·LWC	-65 E-5	42
8	LWC·WST	-38 E-4	1639
9	LWC·CT	-.016	664
10	LWC·CH	68 E-4	275
11	LWC/CT	-.11	4.6
12	WST·WST	.13	754
13	CT·CH	.011	512
14	WV·WV	0	0
15	LWC·LWC·CT	64 E-6	269
16	LWC·LWC·CH	-29 E-6	78

R^2 = .996

σ = 1.62°c

INTERCEPT: 158.9

TABLE 3.3.5

FORWARD MODEL FOR PREDICTING 94 GHz TEMP.

(UNITS GIVEN IN TABLE 3.2.1)

T94

REGRESSION VARIABLES

		COEFF.	F
1	LWC	1.27	771
2	WST	.393	207
3	DLT	.737	110
4	CT	1.43	978
5	CH	.092	16
6	WV	11.6	36
7	LWC·LWC	-54 E-4	598
8	LWC·WST	-15 E-4	62
9	LWC·CT	-.043	1095
10	LWC·CH	-35 E-4	15
11	LWC/CT	-.26	5
12	WST·WST	0	0
13	CT·CH	0	0
14	WV·WV	-2.3	16
15	LWC·LWC·CT	26 E-5	946
16	LWC·LWC·CH	-22 E-6	10

R^2 = .943

σ = 3.4°c

INTERCEPT: 314.0

TABLE 3.3.6

REGRESSION MODEL FOR LWC (milligrams/cm^2)

(defined with 863 points)

variable	coefficient	F value
T19	- 85.4	118.
T37	- 79.8	395.
T94	81.2	126.
1./T19	76.8 E+04	15.6
1./T37	- 12.5 E+05	56.6
1./T94	0.	0.
T19.T37	0.411	1265.
T19.T94	- 0.224	75.7
T37.T94	0.	0.
T19/T37	12.3 E+03	123.
T19/T94	0.	0.
T37/T19	4723	82.6
T37/T94	1273.	23.2
T94/T19	-6342.	127.
T94/T37	0.	0.

dependent mean 88.

R^2 .990

intercept estimate -224.

error mean -2.097 E-03

error standard deviation 9.98

σ_{prior} 99.27

TABLE 3.3,7

REGRESSION MODEL FOR WST (m/sec)

(defined with 863 points)

variable	coefficient	F value
T19	3.66	41.
T37	6.32	220.
T94	- .943	36.7
1./T19	61.4 E+03	42.7
1./T37	0.	0.
1./T94	0.	0.
T19.T37	- 18.4 E-03	135.
T19.T94	0.	0.
T37.T94	0.	0.
T19/T37	-995.	202.
T19/T94	812.	112.
T37/T19	-645.	295.
T37/T94	-936.	104.
T94/T19	0.	0.
T94/T37	0.	0.

dependent mean 13.66

R^2 .848

intercept estimate 442.5

error mean -3.260 E-05

error standard deviation 2.23

σ_{prior} 5.72

TABLE 3.3.8

REGRESSION MODEL FOR CT ($m \times 10^2$)

(defined with 863 data points)

variable	coefficient	F value
T19	- 2.73	59.7
T37	0.	0.
T94	- 1.04	79.5
1./T19	- 34.65 E+03	34.7
1./T37	0.	0.
1./T94	- 22.40 E+03	7.5
T19.T37	65.6 E-04	114.7
T19.T94	0.	0.
T37.T94	0.	0.
T19/T37	0.	0.
T19/T94	0.	0.
T37/T19	0.	0.
T37/T94	0.	0.
T94/T19	0.	0.
T94/T37	141.	81.3

dependent mean 7.75

R^2 .608

intercept estimate 622.23

error mean 2.278 E-04

error standard deviation 4.57

σ_{prior} 7.29

TABLE 3.3.9

REGRESSION MODEL FOR CH ($m \times 10^2$)

(defined with 863 data points)

variable	coefficient	F value
T19	0.	0.
T37	0.	0.
T94	9.07	171.
1./T19	0.	0.
1./T37	0.	0.
1./T94	0.	0.
T19.T37	51.1 E-03	269.
T19.T94	- 29.6 E-03	36.
T37.T94	- 33.7 E-03	388.
T19/T37	3271.	13.
T19/T94	-6950.	41.
T37/T19	-3124.	40.
T37/T94	2334.	8.
T94/T19	1733.	17.
T94/T37	-3910.	22.

dependent mean 33.19

R^2 .730

intercept estimate 5071.

error mean -6.613 E-04

error standard deviation 7.37

σ_{prior} 14.17

TABLE 3.3.10

REGRESSION MODEL FOR WV (gm/cm^2)

(defined with 863 data points)

variable	coefficient	F value
T19	0.	0.
T37	0.	0.
T94	0.	0.
1./T19	-5508.	122.
1./T37	8483.	130.
1./T94	-4629.	53.
T19.T37	- 75.5 E-06	9.
T19.T94	0.	0.
T37.T94	0.	0.
T19/T37	0.	0.
T19/T94	20.8	25.
T37/T19	32.4	98.
T37/T94	0.	0.
T94/T19	0.	0.
T94/T37	0.	0.

dependent mean 1.74

R^2 .304

intercept estimate -39.3

error mean -7.97 E-07

error standard deviation .327

σ_{prior} .390

Table 3.3.11

INVERSION ERRORS USING REGRESSION

(units given in Table 3.2.1)

			LWC	WST	DLT	CT	CH	WV
		a priori mean	64.558	13.639	-.249	6.456	32.010	1.729
		a priori r.m.s.	46.922	5.701	1.918	5.759	13.988	.385
Linear regression		error mean	7.104	-.53	.054	.315	-.242	-.001
		error r.m.s.	38.46	3.53	1.809	5.386	10.75	.349
Non-linear regression	a'	error mean	.096	-.035	.004	-.012	-.059	-.004
		error r.m.s.	6.07	1.99	1.63	4.72	7.60	.316
	b'	error mean	.049	-.034	.006	-.054	.032	-.003
		error r.m.s.	3.97	2.022	1.663	4.39	7.65	0.317

a' ∿ 11/15 terms

b' ∿ 16/21 terms

- noise-free data
- 803 data points for evaluation
- 863 data points for fitting

Table 3.3.12

INVERSION ERRORS USING KALMAN FILTER

(units given in Table 3.2.1)

			LWC	WST	DLT	CT	CH	WV
Extended Kalman Filter	σ_r = .5	error mean	42.96	-5.67	2.06	-.59	-24.20	.097
		error r.m.s.	47.73	3.71	2.72	5.24	25.45	.391
	σ_r = 2.	error mean	34.70	-4.40	1.83	-.62	-21.72	.102
		error r.m.s.	42.03	3.07	2.56	5.24	23.21	.388
Iterated Extended Kalman Filter	σ_r = .5	error mean	-3.65	-1.75	-.181	-.090	2.26	.020
		error r.m.s.	15.93	2.61	1.795	5.397	10.96	.342
	σ_r = 1.	error mean	-4.04	-1.74	-.189	-.143	2.14	.014
		error r.m.s.	13.77	2.62	1.793	5.121	10.62	.335
	σ_r = 2.	error mean	-2.80	-1.68	-.141	-.036	1.45	.009
		error r.m.s.	17.73	2.63	1.799	5.252	10.19	.340
	$\begin{bmatrix} 1 & & \\ & 4 & \\ & & 9 \end{bmatrix}$	error mean	-1.89	-1.75	-.109	-.099	1.08	.0158
		error r.m.s.	16.81	2.57	1.758	5.233	10.25	.337

noise-free data, 803 data points
σ_r = measurement rms noise in model

$$V = \sigma_r^2 \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix} \quad \text{or} \quad V = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 4 & 0 \\ 0 & 0 & 9 \end{bmatrix}$$

Note that using $\sigma_r = 1^0$ minimizes the LWC error. We remark that setting $\sigma_r = 0$ does not minimize the rms error since the measurements are nonlinear. Strictly speaking, the use of $\sigma_r = 0$ with noise-free measurements will yield minimum rms errors only for linear measurements. The errors for the other parameters are almost identical for $\sigma_r = 0.5°$, 1°, 2° and for the case $\sigma_{19} = 1°$, $\sigma_{37} = 2°$, $\sigma_{94} = 3°$. The effect of adding actual noise to the data is shown in Table 3.3.13, where σ_{n_2} represents the rms noise, in °K, added independently to each of the three channels. The LWC and WST errors are degraded only slightly at $\sigma_{n_2} = 1$, but are degraded significantly at $\sigma_{n_2} = 5$. The errors for the other parameters are degraded less significantly.

Table 3.3.14 shows the effect of noise added to both the fitting set (rms value = σ_{n_1}) and the test set (rms value = σ_{n_2}). The degradation is seen to increase for a higher-order regression model, a general characteristic of the regression approach.

Plots of the estimation error for several selected cases for each of the methods studied are given in Figures 3.3.1 - 3.3.19. The actual estimation errors are plotted vs. LWC in each case. Abbreviations are defined in Table 3.3.2.

3.4 Inversions for Data Set #2

3.4.1 Data Base

Several new cases were added to the data base in order to include additional parameters of interest. The 10.69 GHz channel was added to the microwave sensor complement to assess its capability in resolving surface

Table 3.3.13

EFFECT OF ACTUAL MEASUREMENT NOISE ON IEKF

σ_{n_2} = rms measurement noise (°C), 803 points

			LWC	WST	DLT	CT	CH	WV
$\sigma_{n_2} = 1$	$\sigma_r = .5$	error mean	-3.55	-1.79	-.158	-.094	2.01	.027
		error r.m.s.	16.24	2.76	1.818	5.299	11.13	.353
	$\sigma_r = 1.$	error mean	-3.47	-1.80	-.159	-.120	1.83	.029
		error r.m.s.	15.56	2.75	1.798	5.194	10.78	.346
	$\sigma_r = 2.$	error mean	-3.31	-1.69	-.149	-.027	1.58	.011
		error r.m.s.	18.34	2.74	1.809	5.235	10.50	.344
	$\begin{bmatrix} 1 & & \\ & 4 & \\ & & 9 \end{bmatrix}$	error mean	-2.54	-1.72	-.124	-.122	1.19	.022
		error r.m.s.	16.83	2.72	1.767	5.174	10.50	.341
$\sigma_{n_2} = 5$	$\sigma_r = .5$	error mean	-6.71	-2.03	-.223	-.555	2.29	.048
		error r.m.s.	35.90	4.63	2.013	7.268	16.13	.453
	$\sigma_r = 1.$	error mean	-7.10	-1.84	-.246	-.584	2.38	.032
		error r.m.s.	33.87	4.61	1.999	6.832	15.33	.435
	$\sigma_r = 2.$	error mean	-7.87	-1.70	-.250	-.302	2.81	.010
		error r.m.s.	32.59	4.22	1.951	6.167	13.53	.399
	$\begin{bmatrix} 1 & & \\ & 4 & \\ & & 9 \end{bmatrix}$	error mean	-6.89	-1.94	-.223	-.362	2.40	.020
		error r.m.s.	32.69	4.43	1.951	6.237	13.79	.405

Table 3.3.14

EFFECT OF NOISE ON NONLINEAR REGRESSIONS

σ_{n_1} = rms noise in training set

σ_{n_2} = rms noise in test set

		LWC	WST	DLT	CT	CH	WV
$\sigma_{n_1} = 0$ $\sigma_{n_2} = 0$	error mean	.096	-.035	.004	-.012	-.059	-.004
	error r.m.s.	6.07	1.99	1.63	4.72	7.60	.316
	# terms	*11/15	9/15	6/15	6/15	10/15	6/15
$\sigma_{n_1} = 0$ $\sigma_{n_2} = 1$	error mean	-.082	.105	.011	.034	-.057	.008
	error r.m.s.	7.17	2.235	1.64	4.75	8.25	.324
	# terms	11/15	9/15	6/15	6/15	10/15	6/15
$\sigma_{n_1} = 1$ $\sigma_{n_2} = 1$	error mean	-.056	.073	.019	.035	-.295	.009
	error r.m.s.	6.19	2.22	1.65	4.75	8.14	.323
	# terms	14/15	9/15	6/15	6/15	9/15	6/15
$\sigma_{n_1} = 0$ $\sigma_{n_2} = 0$	error mean	.049	-.034	.006	-.054	.032	-.003
	error r.m.s.	3.97	2.022	1.663	4.39	7.65	0.317
	# terms	15/21	9/22	5/22	8/21	8/22	6/20
$\sigma_{n_1} = 0$ $\sigma_{n_2} = 1$	error mean	-.134					
	error r.m.s.	6.18					
	# terms	15/21					
$\sigma_{n_1} = 1$ $\sigma_{n_2} = 1$	error mean	-.065					
	error r.m.s.	6.65					
	# terms	10/21					

*a/b : a = no. of terms in regression model
b = total no. of terms considered

Figure 3.3.1

LWC ERROR

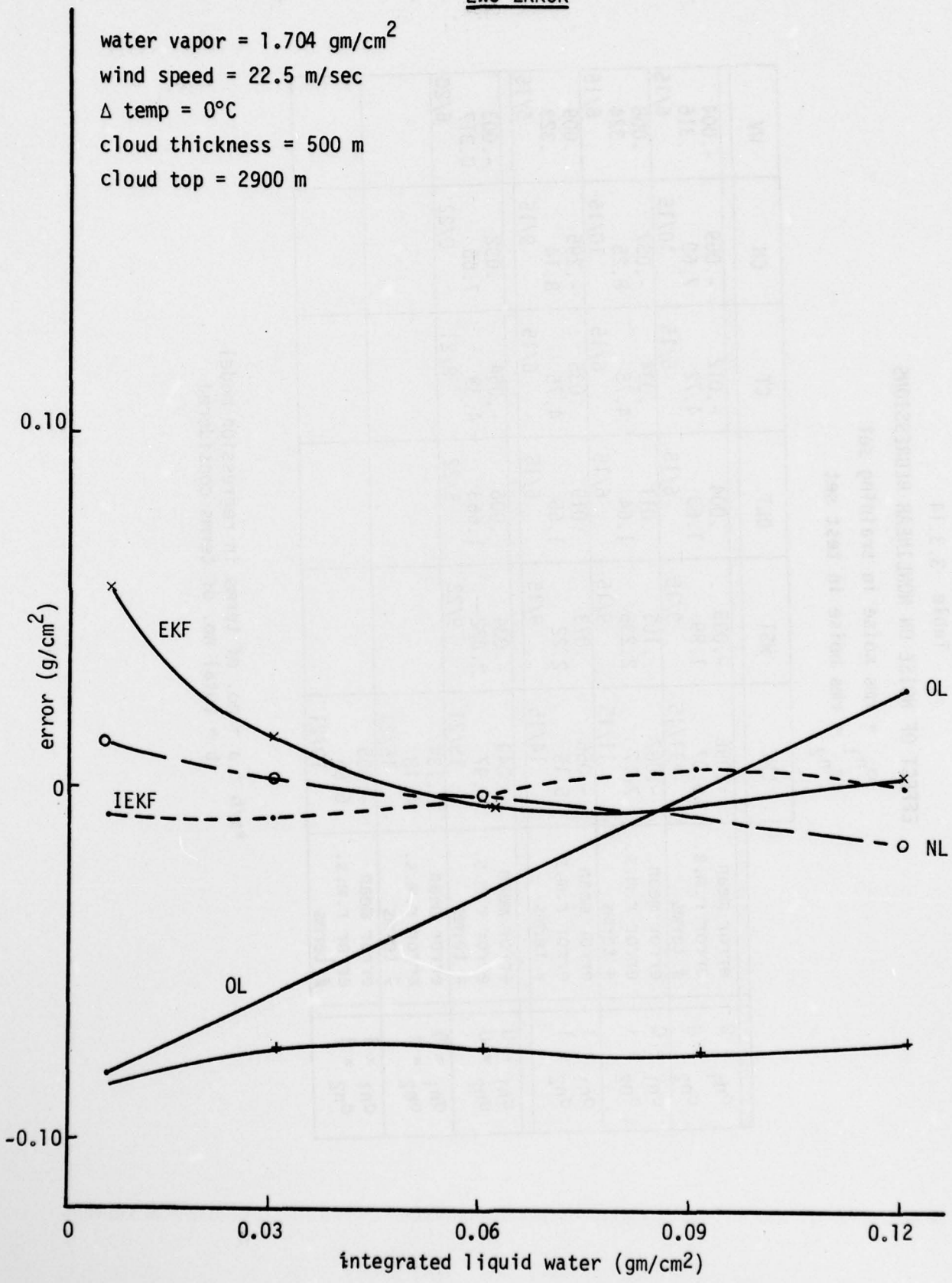


Figure 3.3.2

LWC ERROR

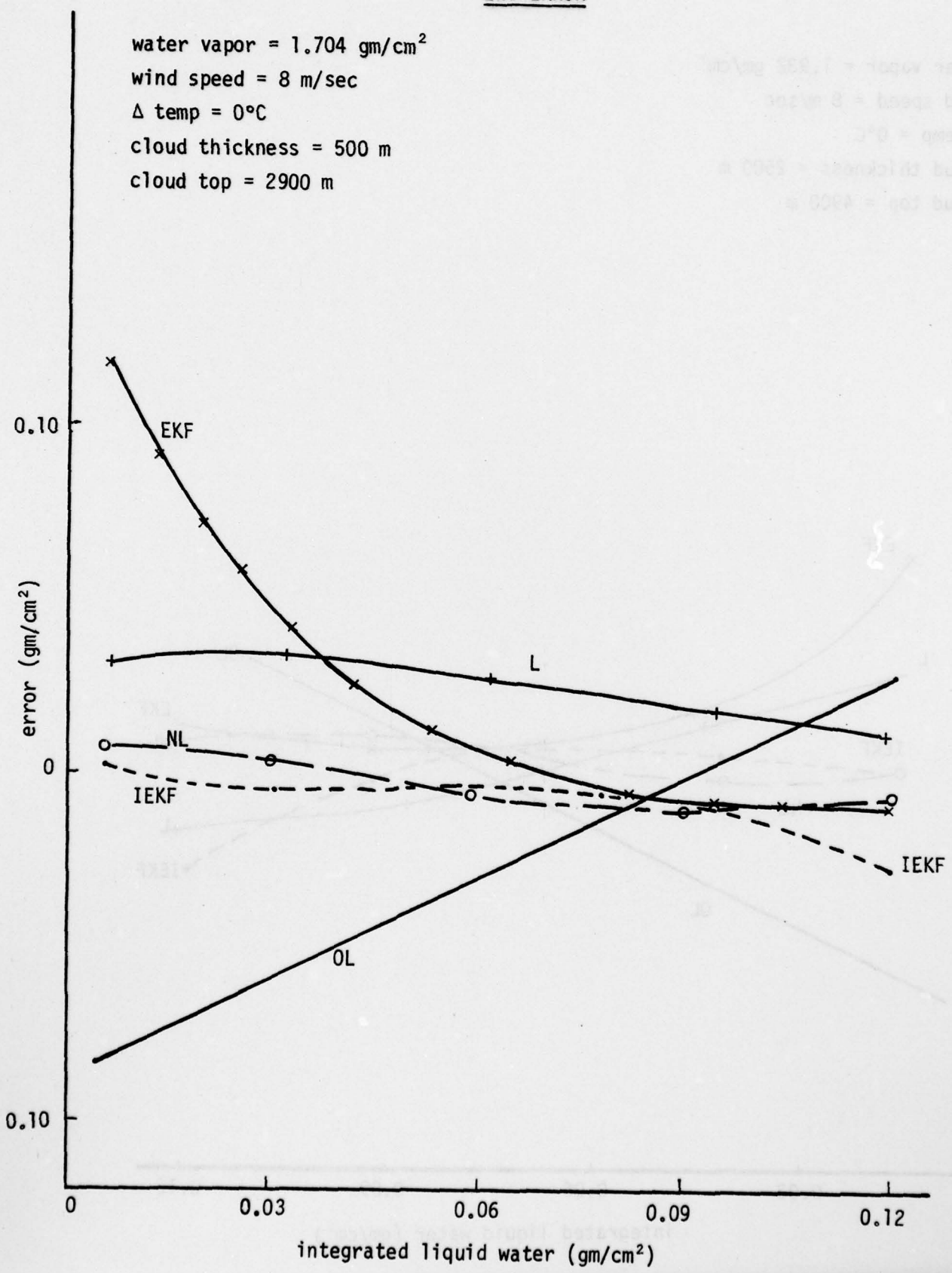


Figure 3.3.3

LWC ERROR

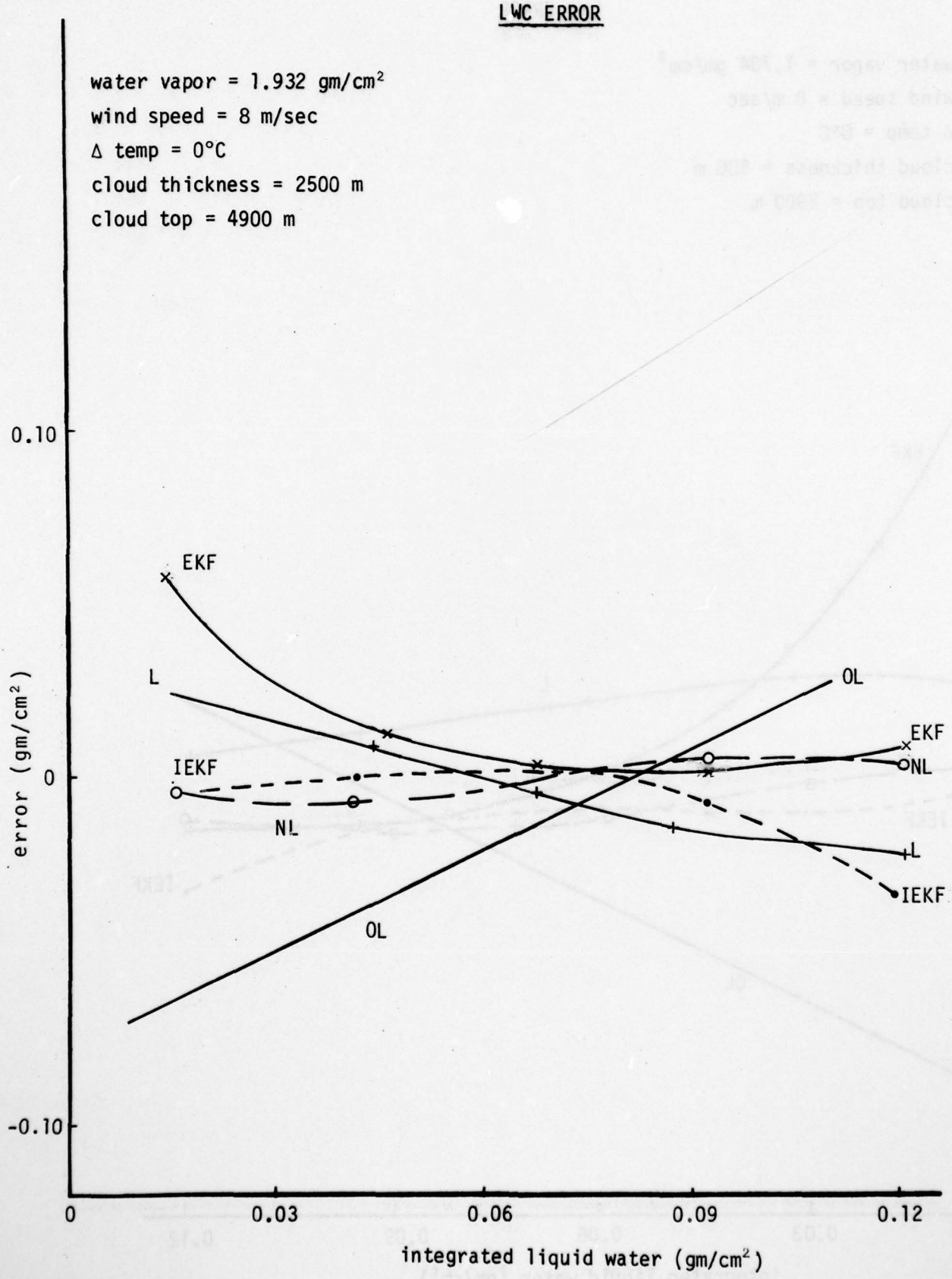


Figure 3.3.4

LWC ERROR

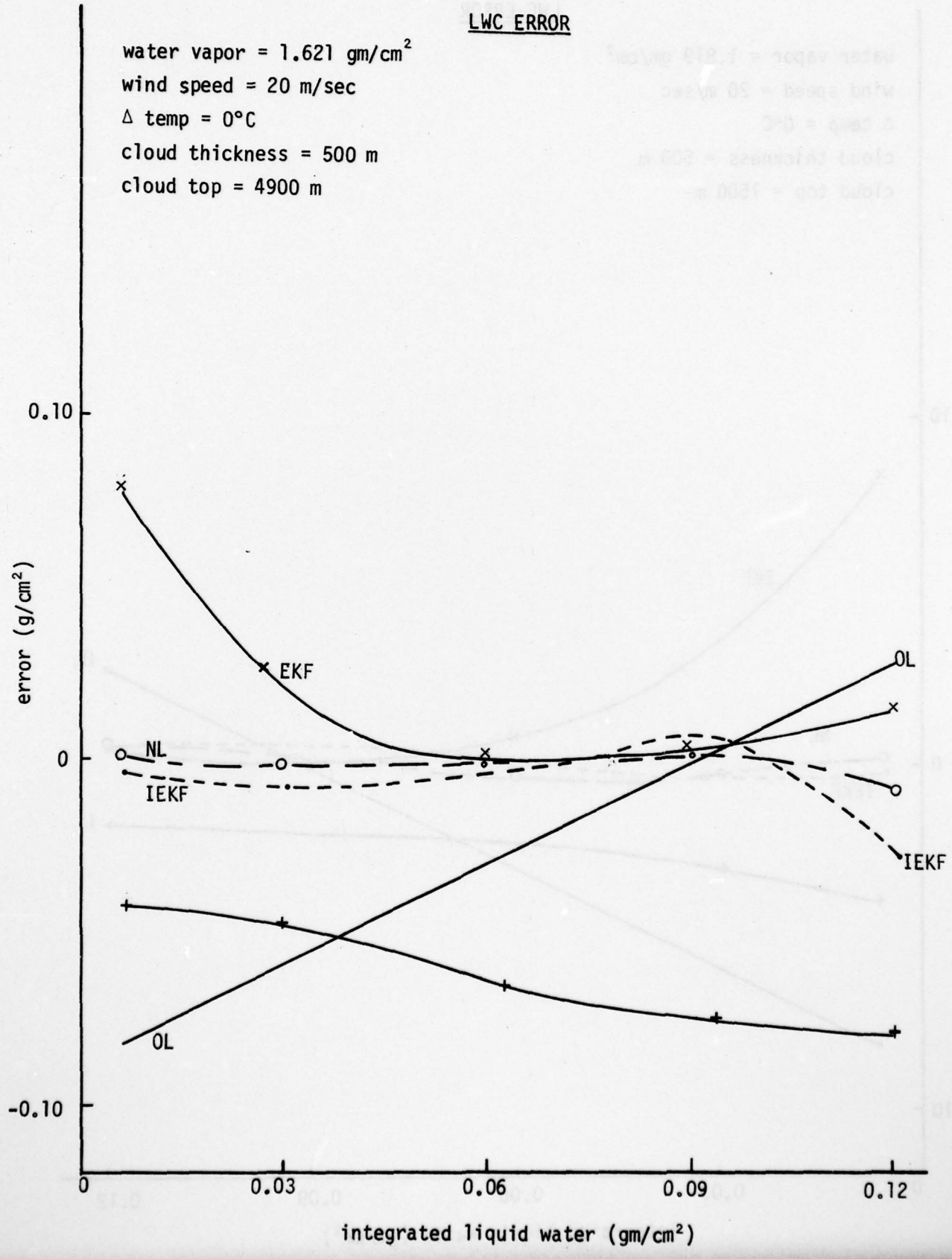


Figure 3.3.5

LWC ERROR

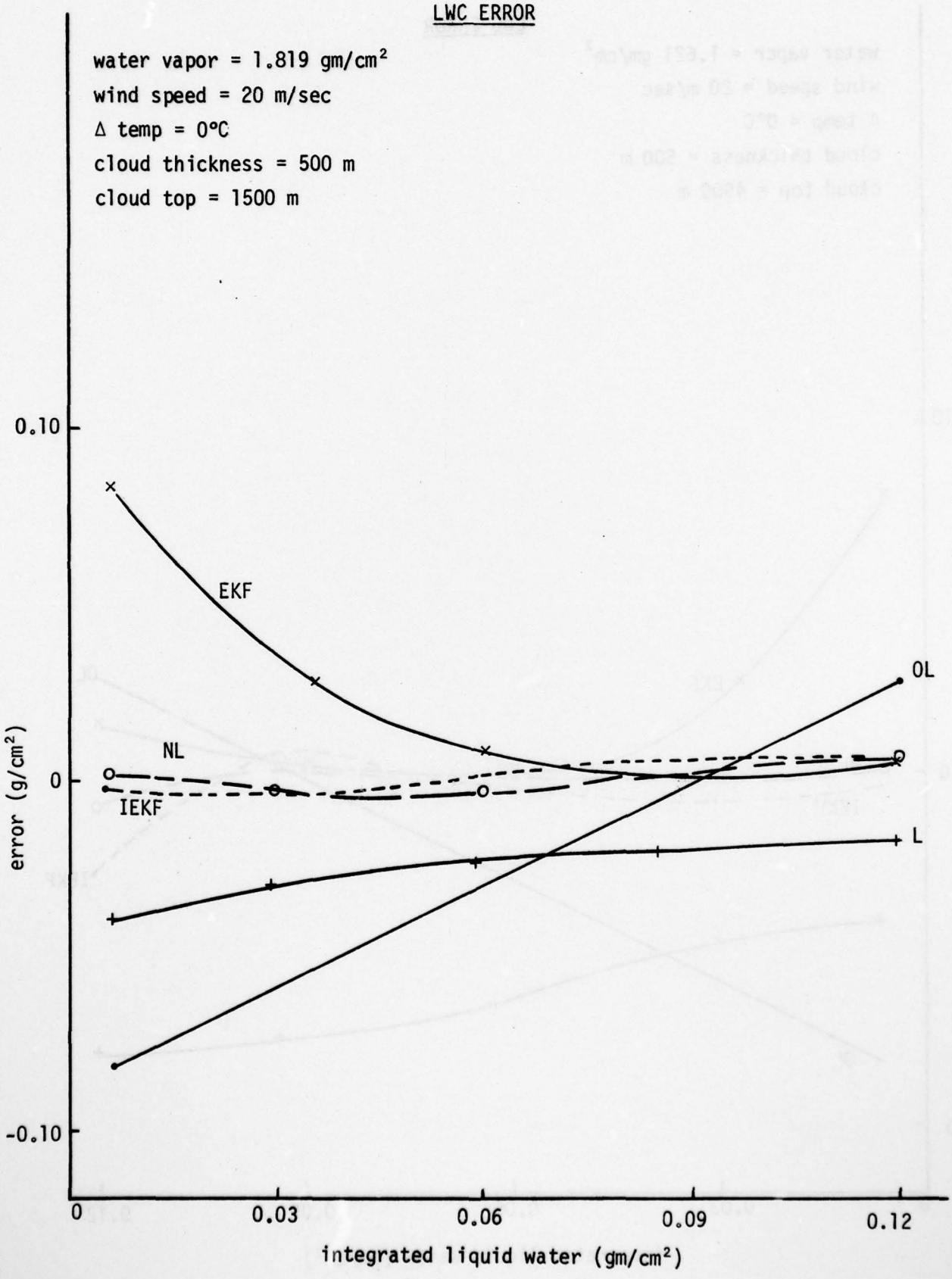


Figure 3.3.6

WST ERROR

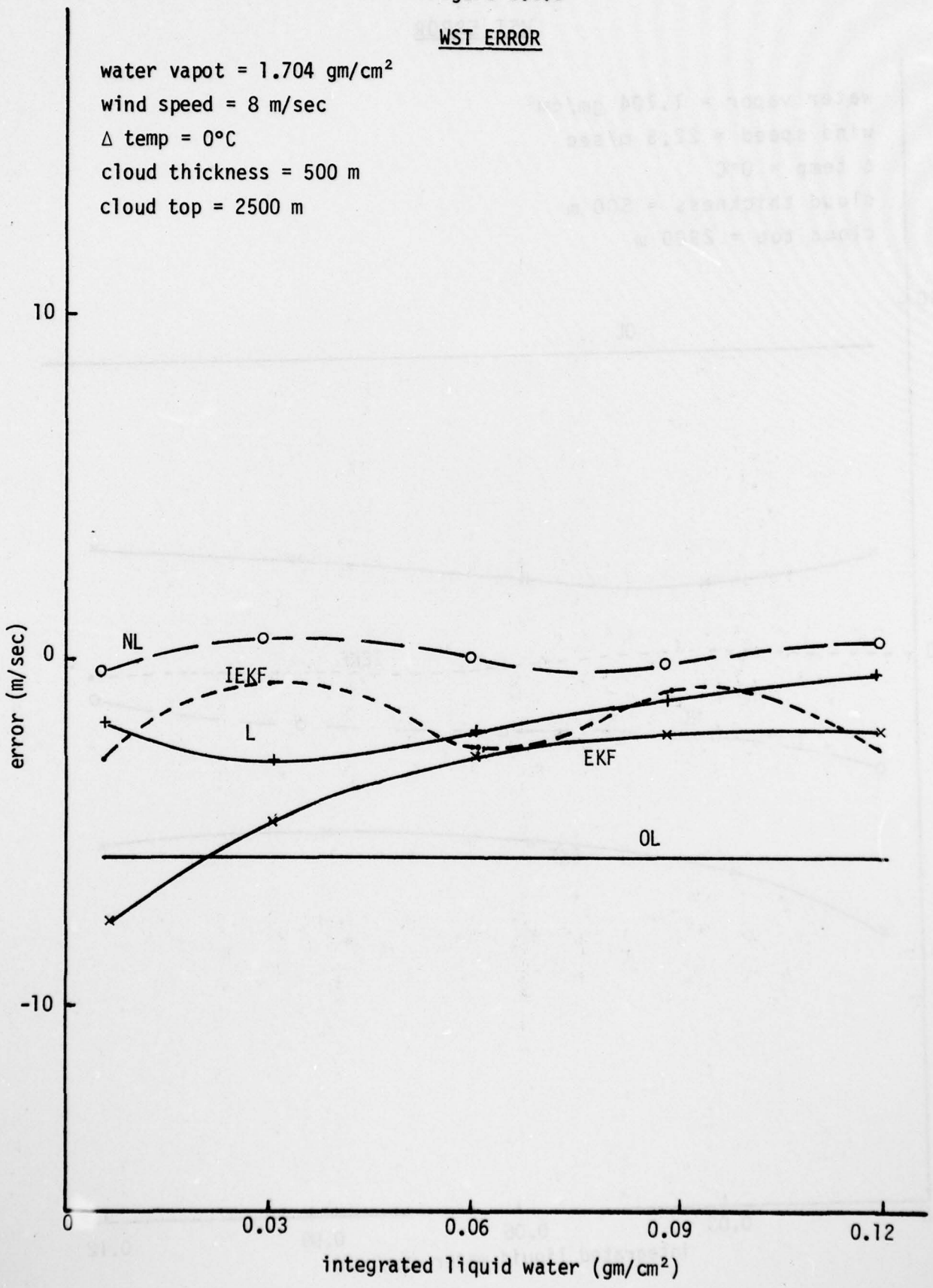


Figure 3.3.7

WST ERROR

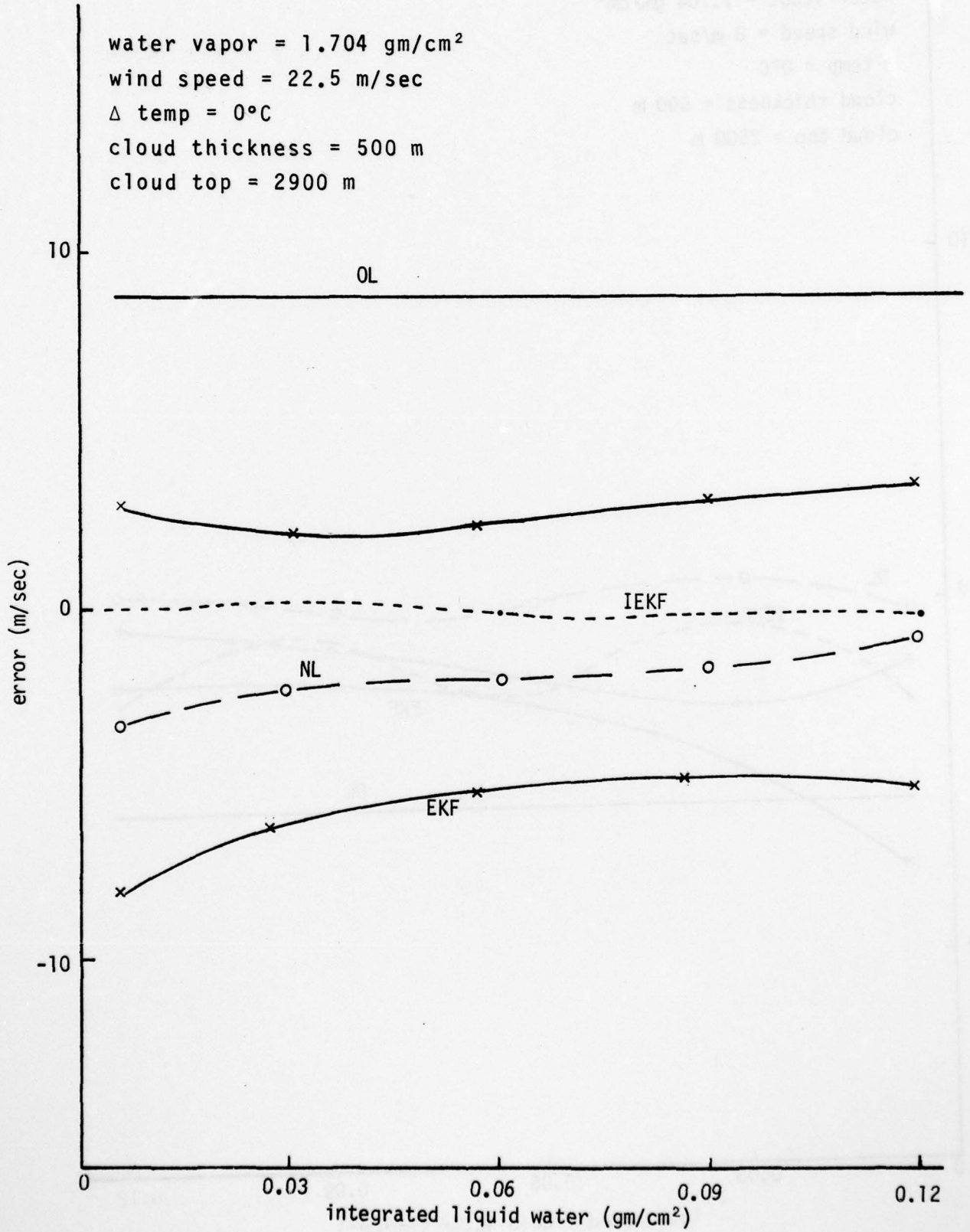
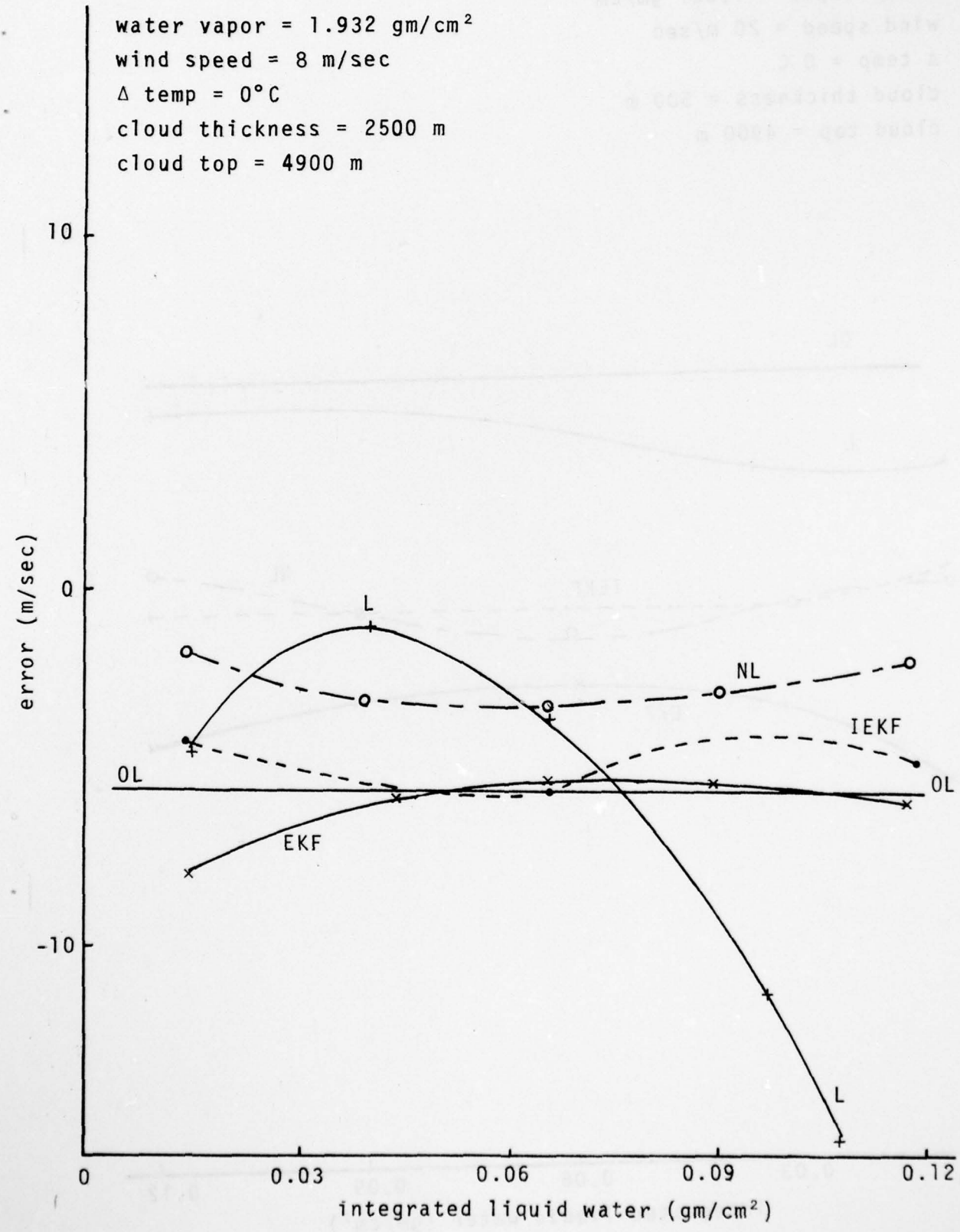


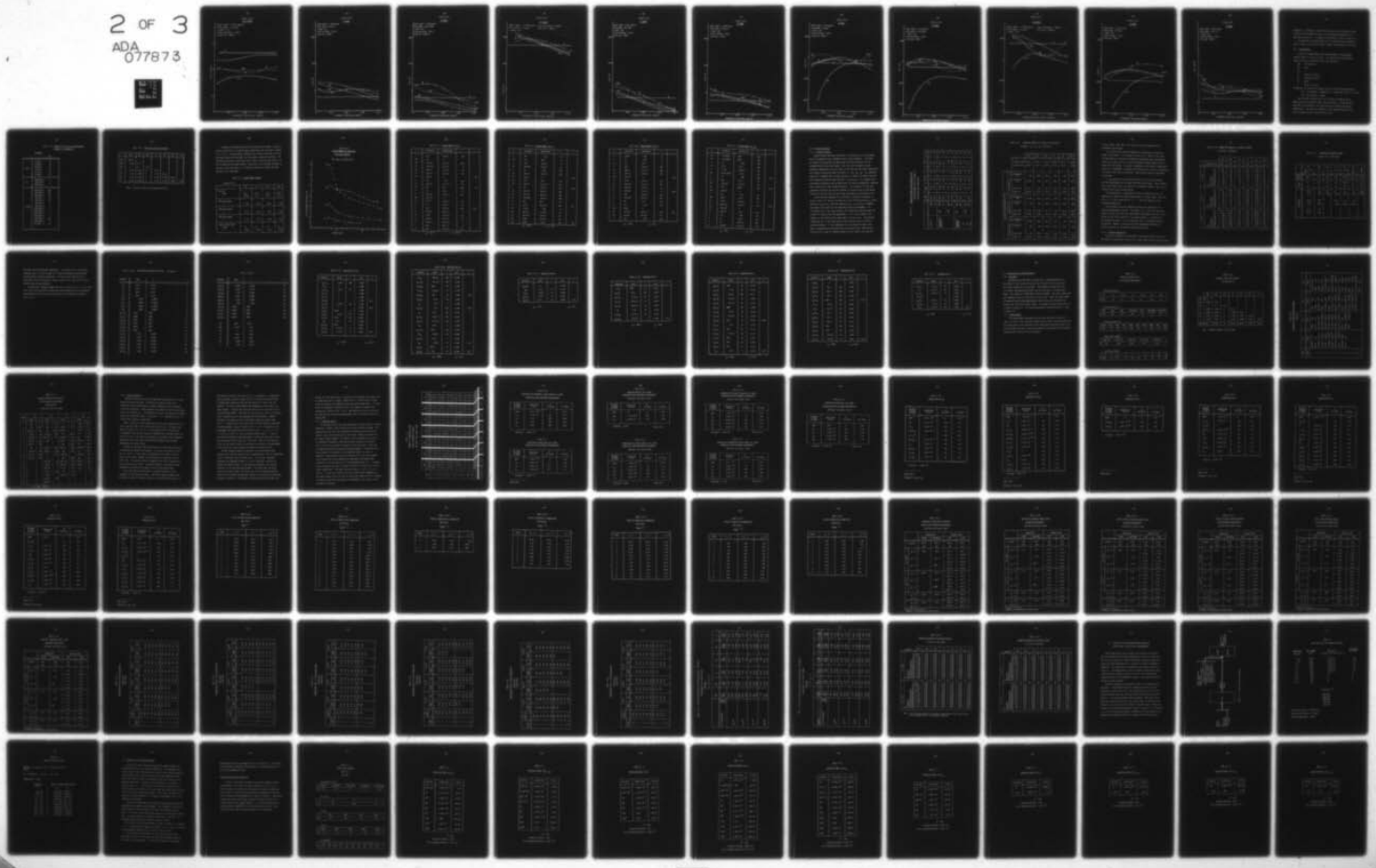
Figure 3.3.8

WST ERROR



AD-A077 873 SCIENTIFIC SYSTEMS INC CAMBRIDGE MA F/G 4/2
MULTISPECTRAL CLOUD IDENTIFICATION STUDY.(U)
SEP 78 D E GUSTAFSON , W H LEDSHAM F19628-77-C-0208
UNCLASSIFIED AFGL-TR-78-0280 NL

2 OF 3
ADA
077873



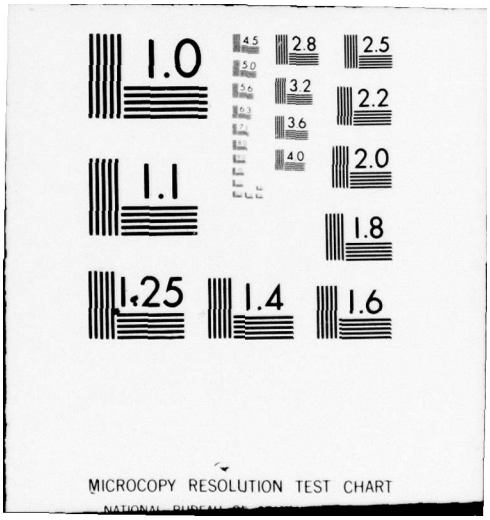

1.0
1.1
1.25
1.4
1.6
1.8
2.0
2.2
2.5
2.8
3.2
3.6
4.0
4.5
5.0
5.6
6.3
7.1
MICROCOPY RESOLUTION TEST CHART

Figure 3.3.9

WST ERROR

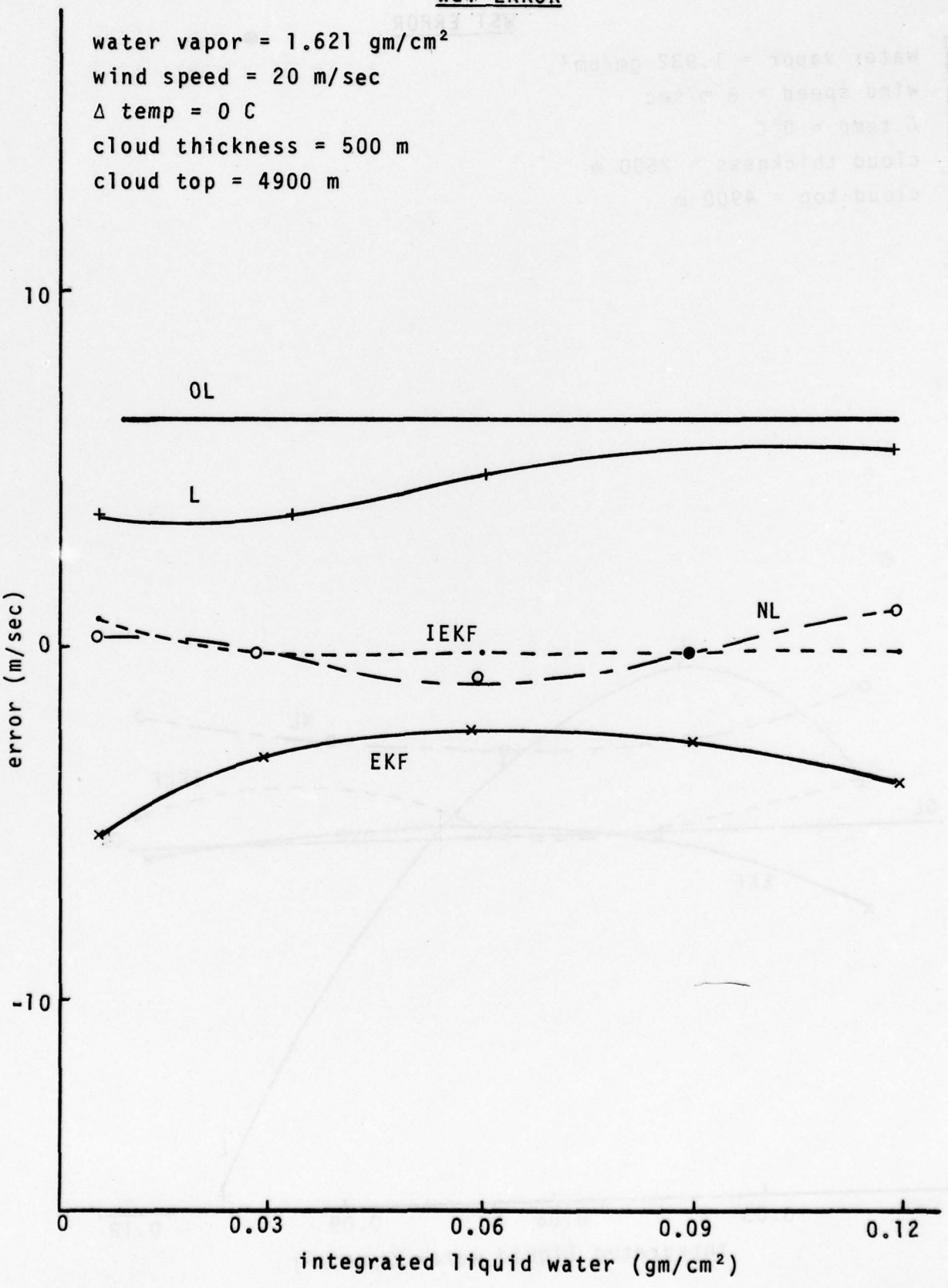


Figure 3.3.10

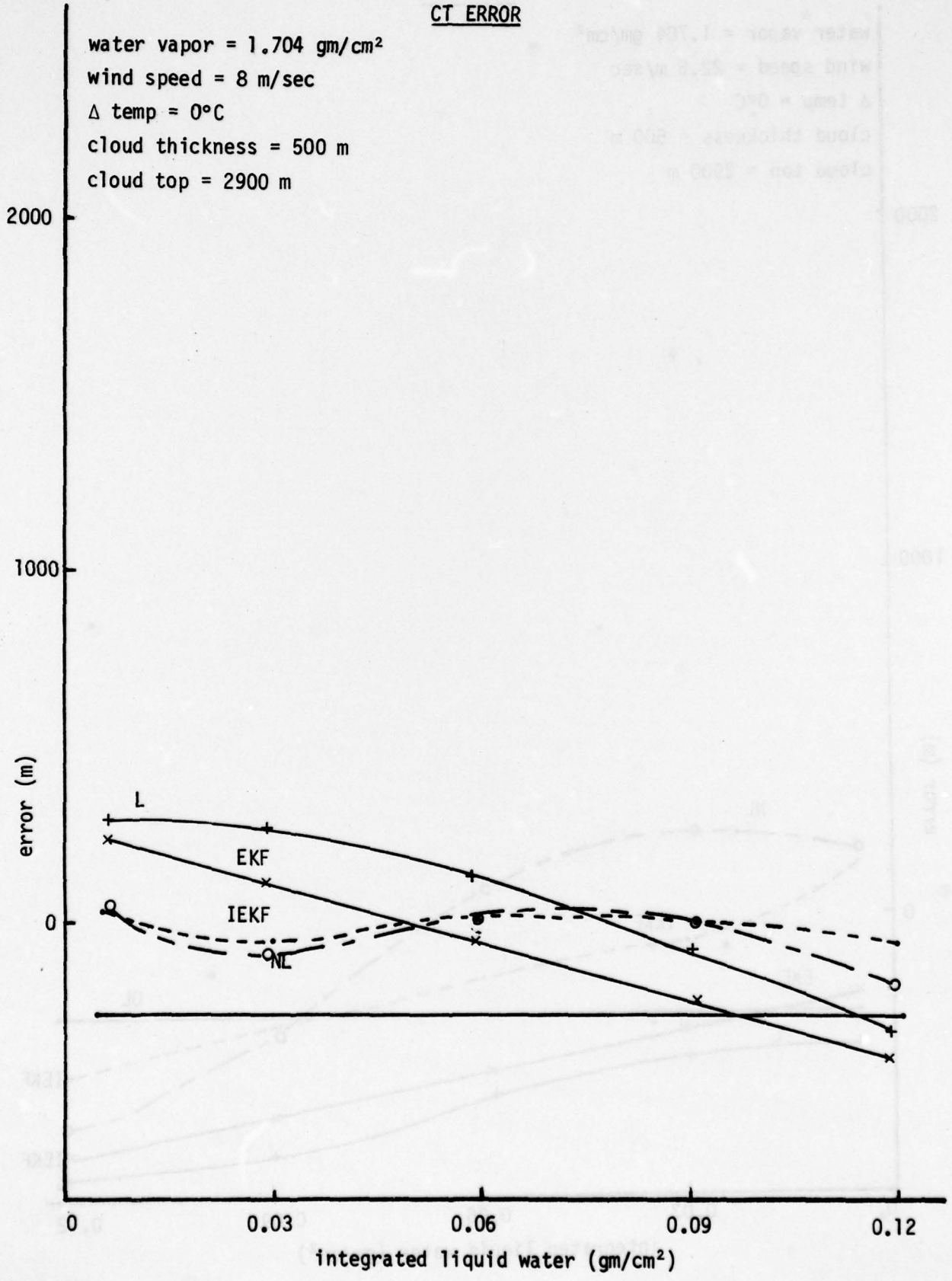

CT ERROR
water vapor = 1.704 gm/cm²
wind speed = 8 m/sec
Δ temp = 0°C
cloud thickness = 500 m
cloud top = 2900 m
2000
1000
0
error (m)
L
EKF
IEKF
NL
0
0.03
0.06
0.09
0.12
integrated liquid water (gm/cm²)

Figure 3.3.11

CT ERROR

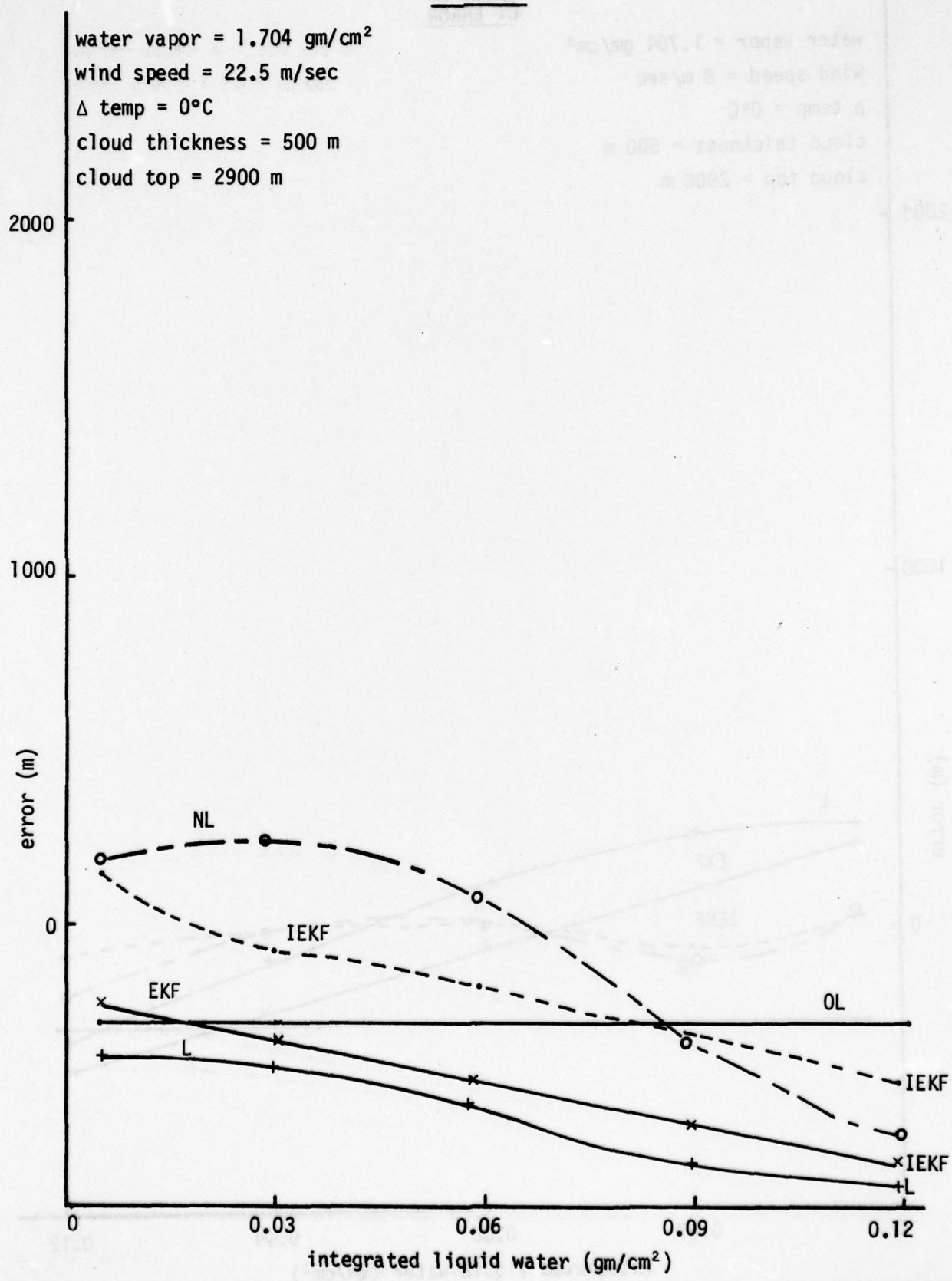


Figure 3.3.12

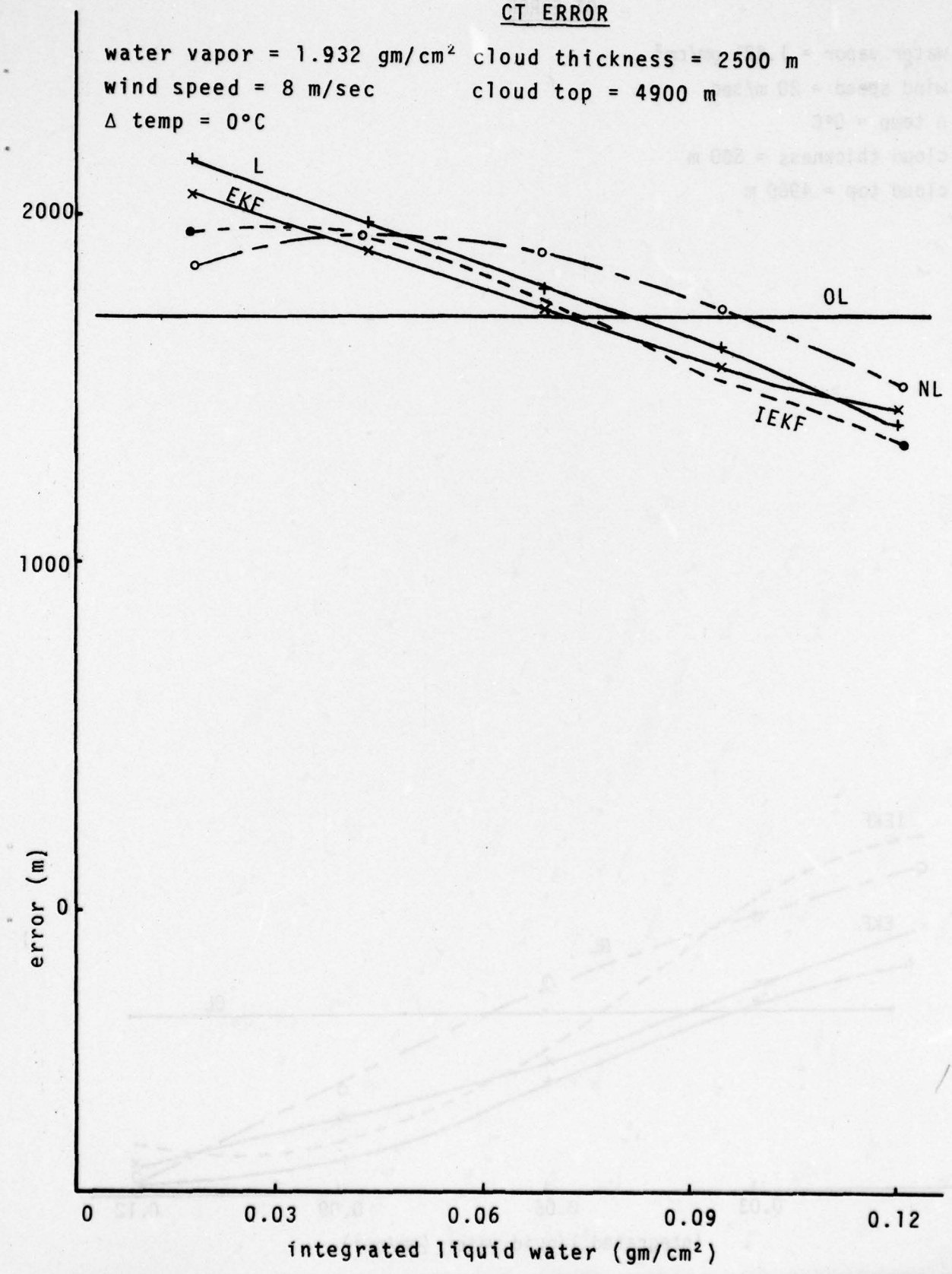

CT ERROR
water vapor = 1.932 gm/cm²
wind speed = 8 m/sec
Δ temp = 0°C
cloud thickness = 2500 m
cloud top = 4900 m
L
EKF
OL
NL
IEKF
2000
1000
0
error (m)
0
0.03
0.06
0.09
0.12
integrated liquid water (gm/cm²)

Figure 3.3.13

CT ERROR

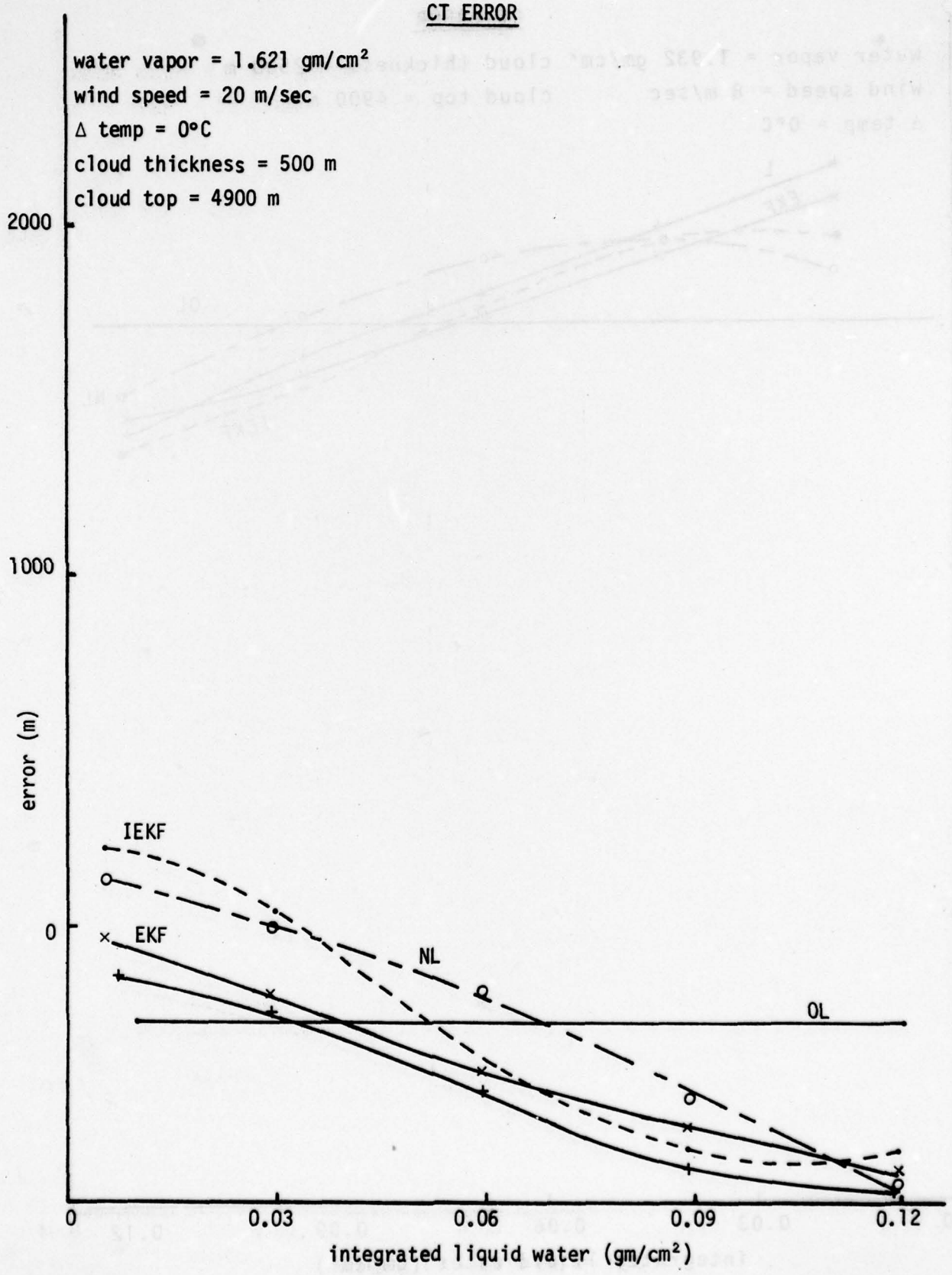


Figure 3.3.14

CH ERROR

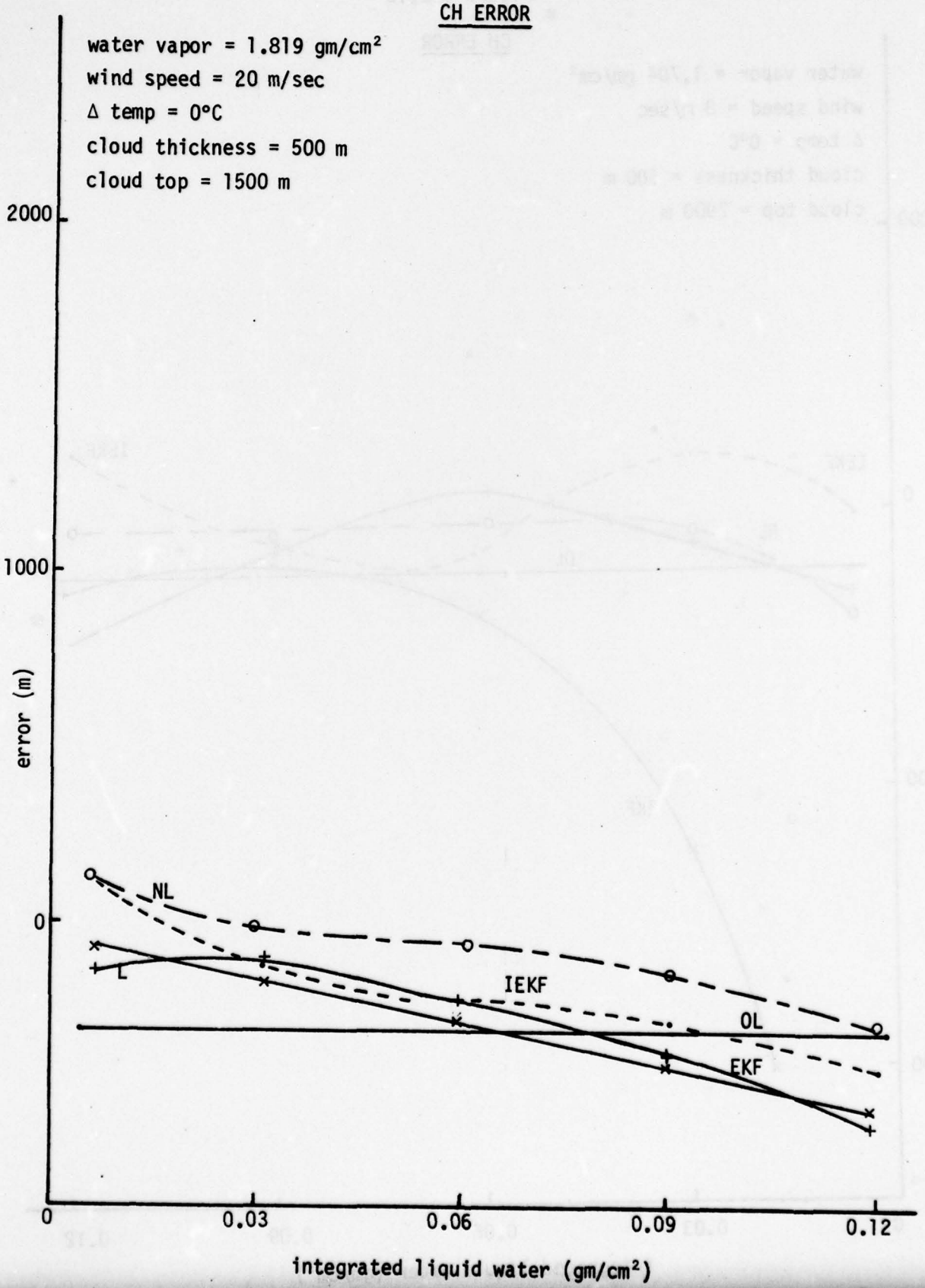


Figure 3.3.15

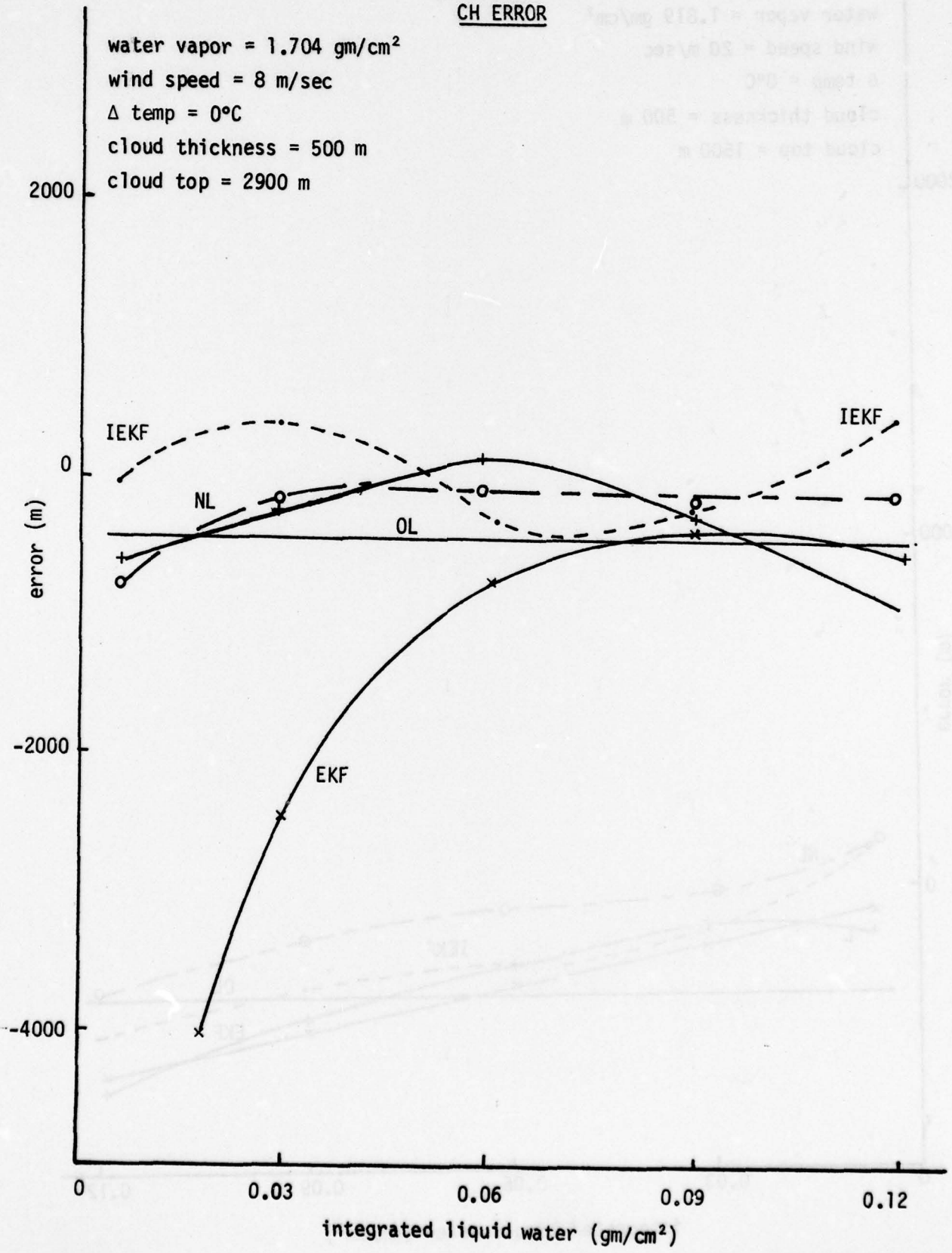

CH ERROR
water vapor = 1.704 gm/cm²
wind speed = 8 m/sec
Δ temp = 0°C
cloud thickness = 500 m
cloud top = 2900 m
2000
0
-2000
-4000
error (m)
IEKF
IEKF
NL
OL
EKF
0
0.03
0.06
0.09
0.12
integrated liquid water (gm/cm²)

Figure 3.3.16

CH ERROR

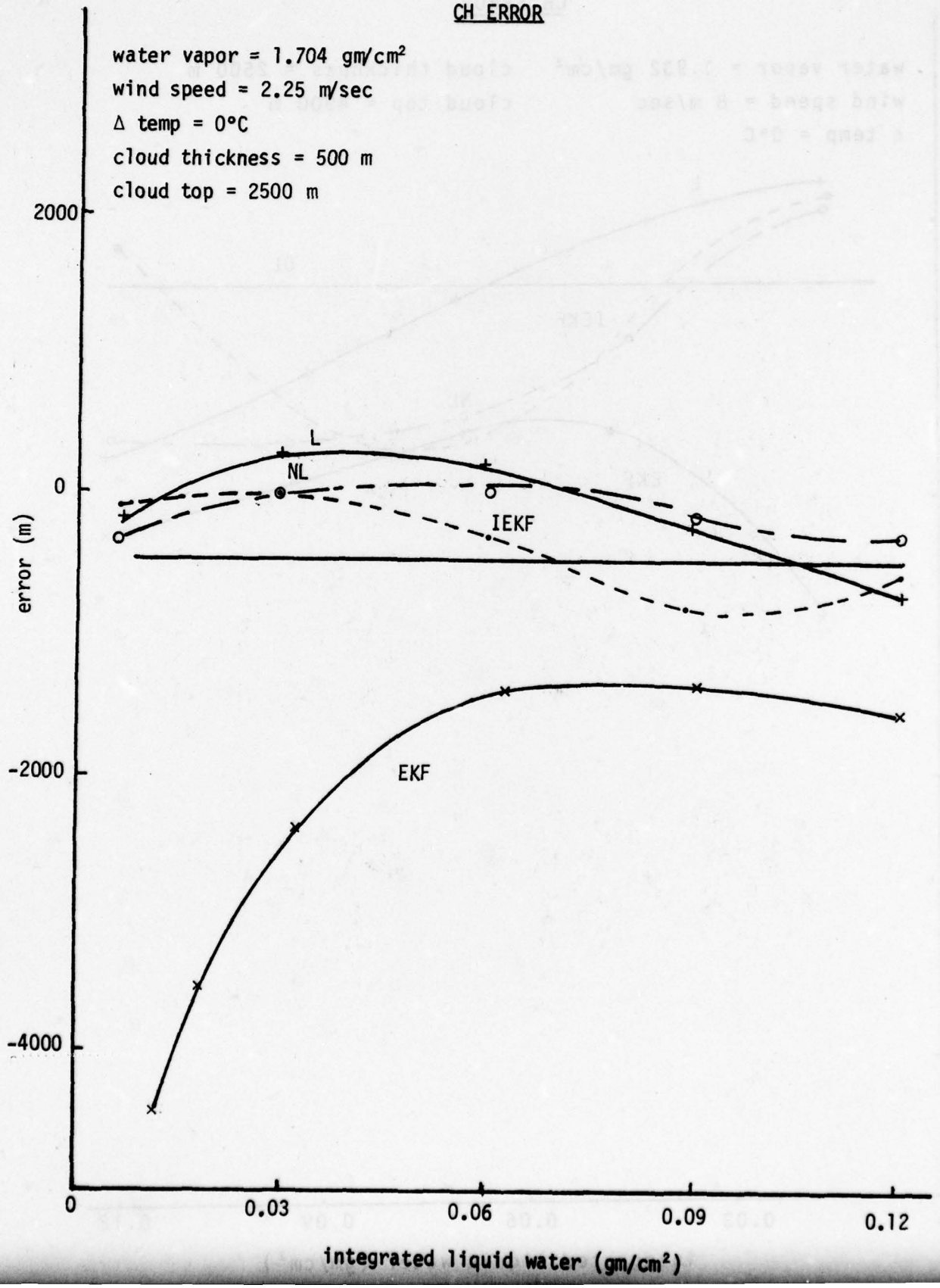


Figure 3.3.17

CH ERROR

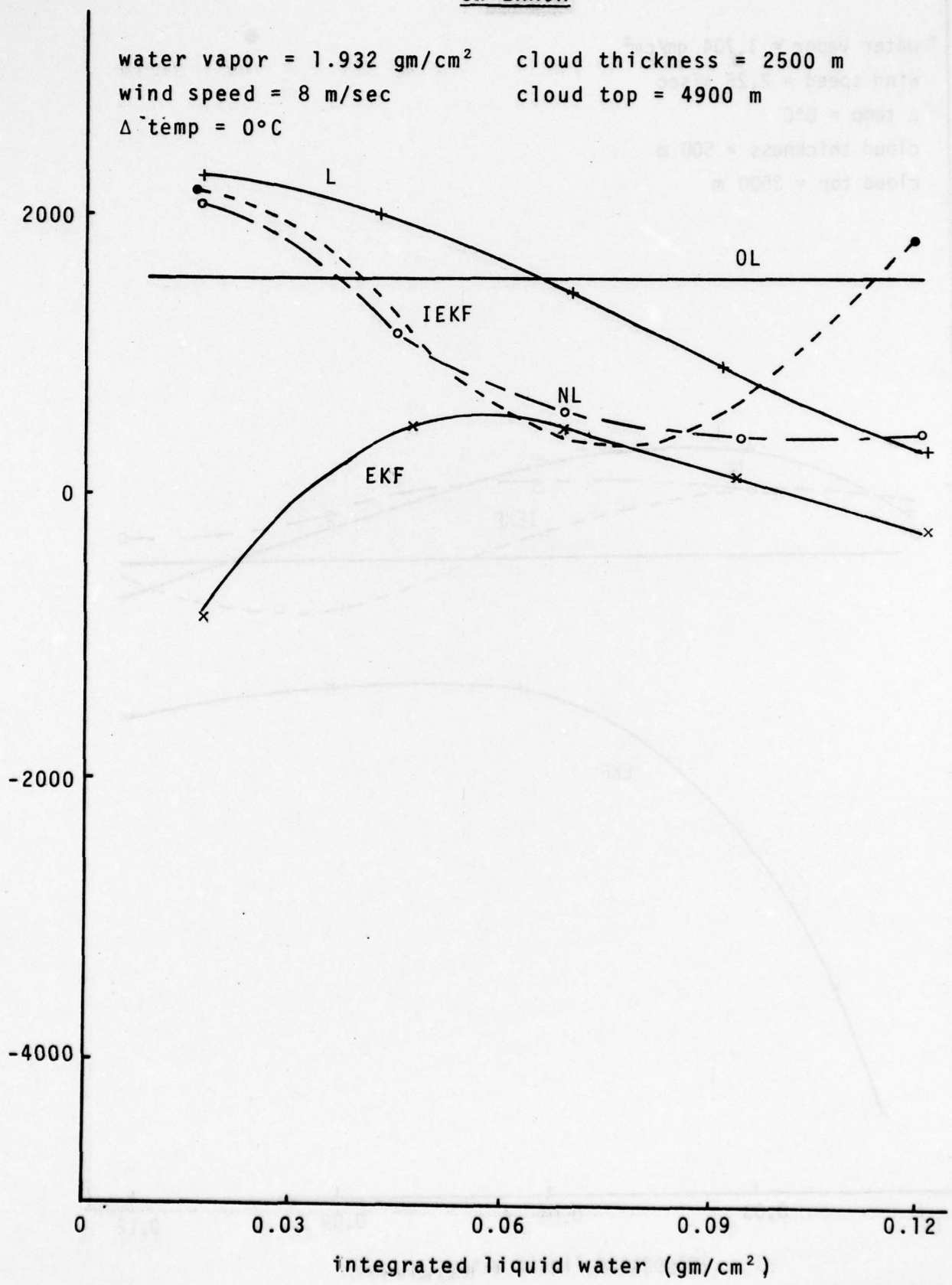


Figure 3.3.18

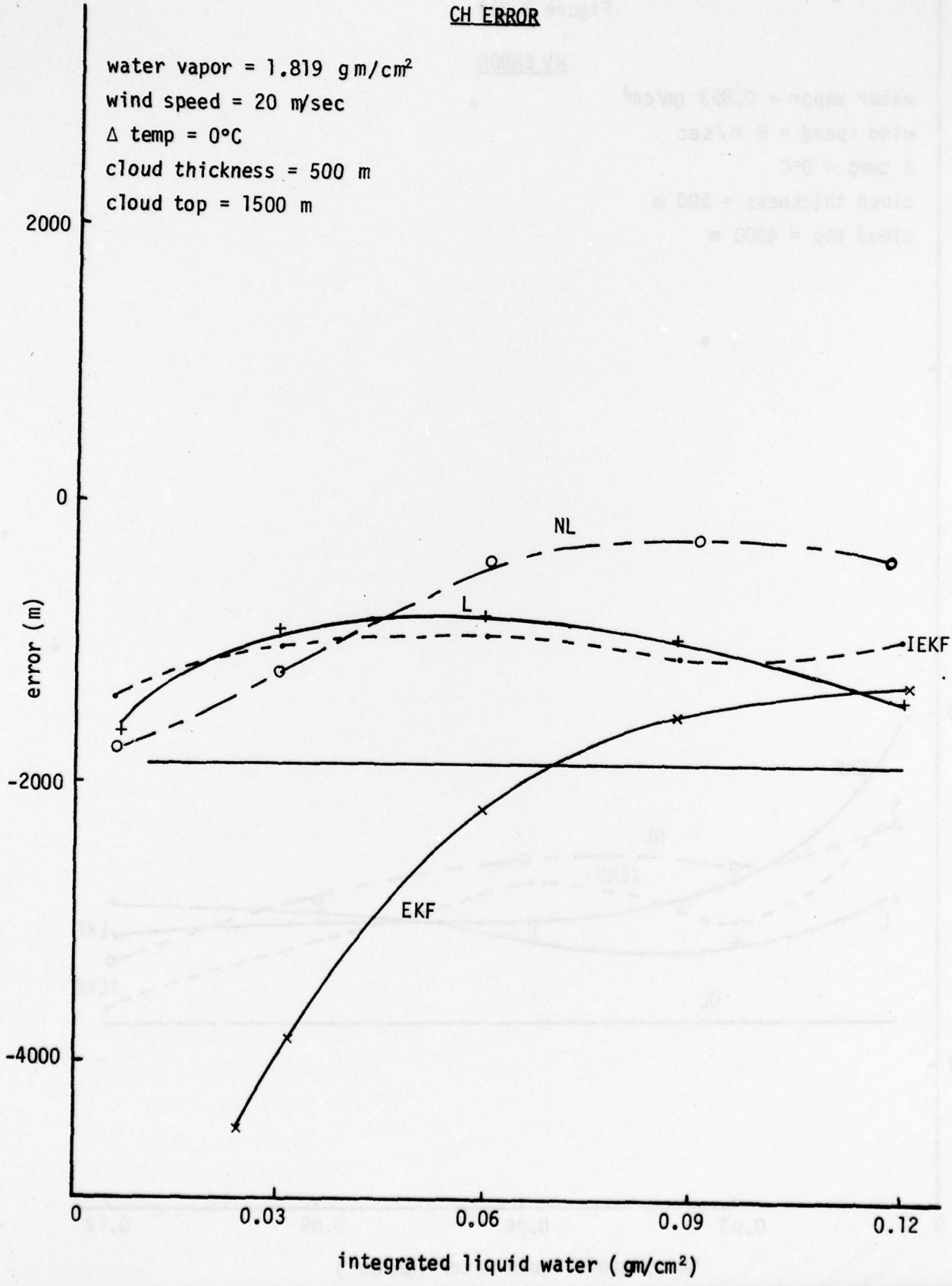
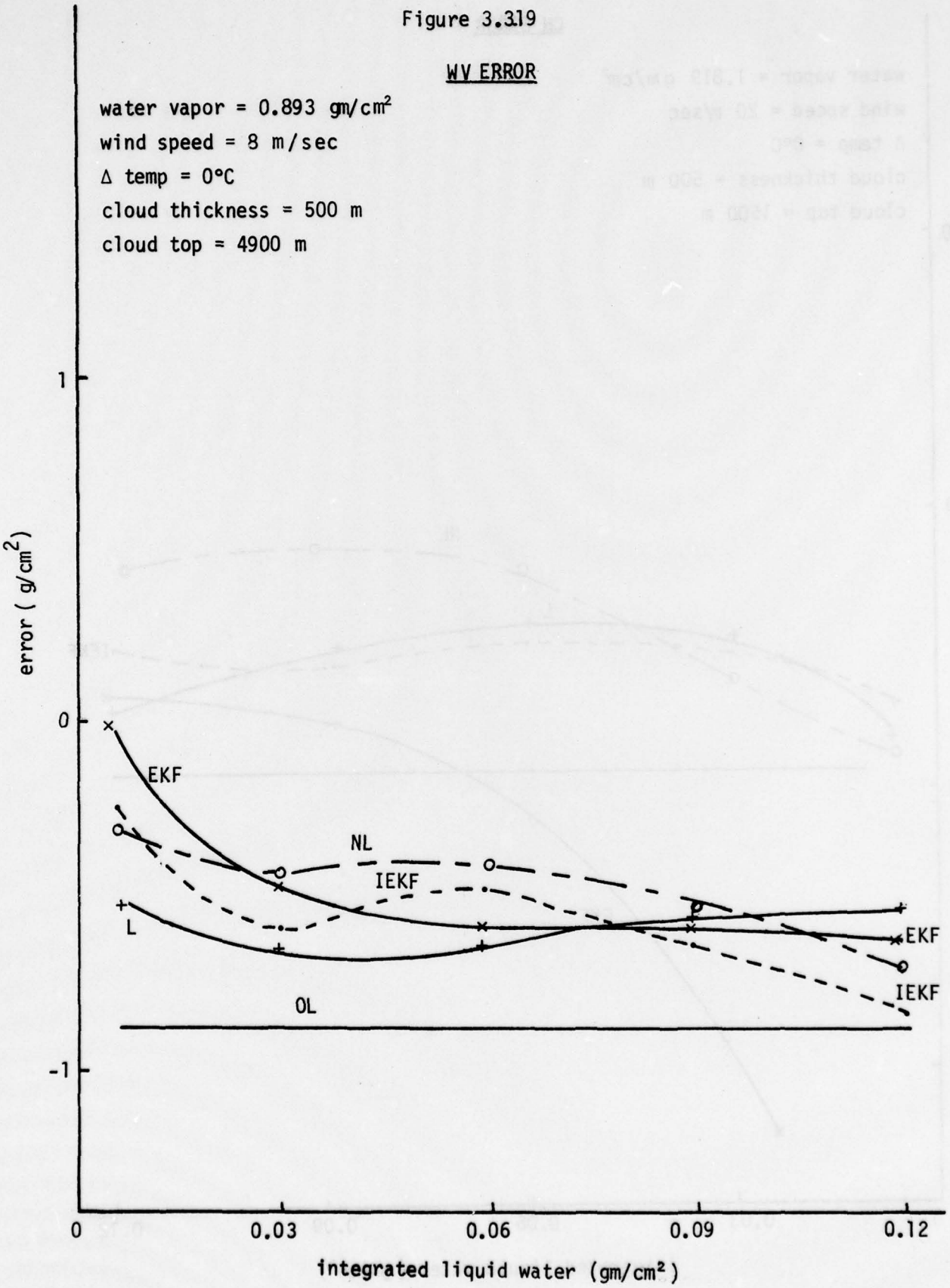

CH ERROR
water vapor = 1.819 g m/cm²
wind speed = 20 m/sec
Δ temp = 0°C
cloud thickness = 500 m
cloud top = 1500 m
2000
0
-2000
-4000
error (m)
NL
L
IEKF
EKF
0
0.03
0.06
0.09
0.12
integrated liquid water (gm/cm²)

Figure 3.319

WV ERROR

water vapor = 0.893 gm/cm^2
wind speed = 8 m/sec
Δ temp = 0°C
cloud thickness = 500 m
cloud top = 4900 m



parameters. In addition, a series of rain cases and ice clouds were added. A total of 941 microwave cases were run using detailed simulations, as described in Section 3.1. A summary of these cases is given in Table 3.4.1.

The predictor correlation matrix for the variables to be estimated is shown in Table 3.4.2 with the diagonal elements corresponding to rms values.

3.4.2 Forward Models

A series of forward models have been generated for the microwave channels based on the 941 test cases. The approach used was nonlinear regression using the following seven basic predictors:

LWC	milligram/cm^2
WST	m/sec
DLT	°K
CT	hundreds of meters
CH	hundreds of meters
WV	gm/cm^2
R	mm/hr (rain rate)

In addition, the following variables related to foam cover were employed:

$$WCC = -0.00188\ WST + 0.000263\ WST^2 - 0.00000\ WST^3 + 18.3\ E - 07\ WST^4$$

$$SCC = WCC(0.255\ WST - 0.299)$$

These variables are empirical fits to Cardone's (1969) white cap model and Ross and Cardone's (1974) total foam model. The two variables represent the percentage of the ocean covered by whitecaps and streaks respectively. They have been presented as two separate variables since their microwave emission characteristics differ.

Table 3.4.1 Summary of Test Cases for Microwave Channels - Data Base #2

MICROWAVE

941

RAIN	3 mm/hr:	12
	15 mm/hr:	16
	30 mm/hr:	16
	24 mm/hr:	22
	12 mm/hr:	22
	24 mm/hr:	22
ICE	5000-6000 m:	4
	5000-7000 m:	2
	5000-8000 m:	2
LIQUID	2400-2900 m:	342
	4400-4900 m:	200
	2400-4900 m:	62
	330-660 m:	40
	660-1320 m:	40
	400-2900 m:	2
	900-2900 m:	2
	1400-2900 m:	2
	2400-4400 m:	2
	2400-3900 m:	2
	1900-2900 m:	2
	1000-1500 m:	100
	2400-3400 m:	2
	NO CLOUD m:	25

Table 3.4.2 Predictor Correlation Matrix

	LWC	WST	DLT	CT	CH	WV	R
LWC	312.25						
WST	0.019	5.75					
DLT	0.	0.	2.21				
CT	0.823	0.027	0.	9.80			
CH	0.157	0.048	0.	0.319	13.55		
WV	0.657	0.009	0.	0.650	-0.026	0.477	
R	0.861	0.016	0.	0.719	0.118	0.599	5.78

Note: diagonal elements are standard deviations

A summary of the modeling errors for different order models is given in Table 3.4.3 and Fig. 3.4.1 and the most significant terms are given in Tables 3.4.4 - 3.4.7. In these tables, predictors are listed in the order in which they were entered into the model via the stepwise regression procedure. The coefficients and the F ratios are given, as well as the rms error after 7, 10 and 15 predictors have been entered. Also given is the intercept estimate y_0 and the a priori rms error σ_0. An asterisk indicates a variable that was deleted in the final model.

Table 3.4.3 Forward Model Summary

• errors in °K

	T_{10}	T_{19}	T_{37}	T_{94}
A priori				
mean	129.1	164.2	202.8	259.2
σ	18.63	31.0	30.0	13.5
$7^{\underline{th}}$ Order Model				
σ	1.00	2.25	6.02	4.01
$10^{\underline{th}}$ Order Model				
σ	0.75	1.56	3.66	3.64
$15^{\underline{th}}$ Order Model				
σ	0.61	1.15	2.87	3.52
Highest Order Model				
order	19	27	27	24
σ	0.59	0.94	2.05	2.63

Figure 3.4.1

Forward Model Fit Error for

Microwave Channels

941 cases, including rain

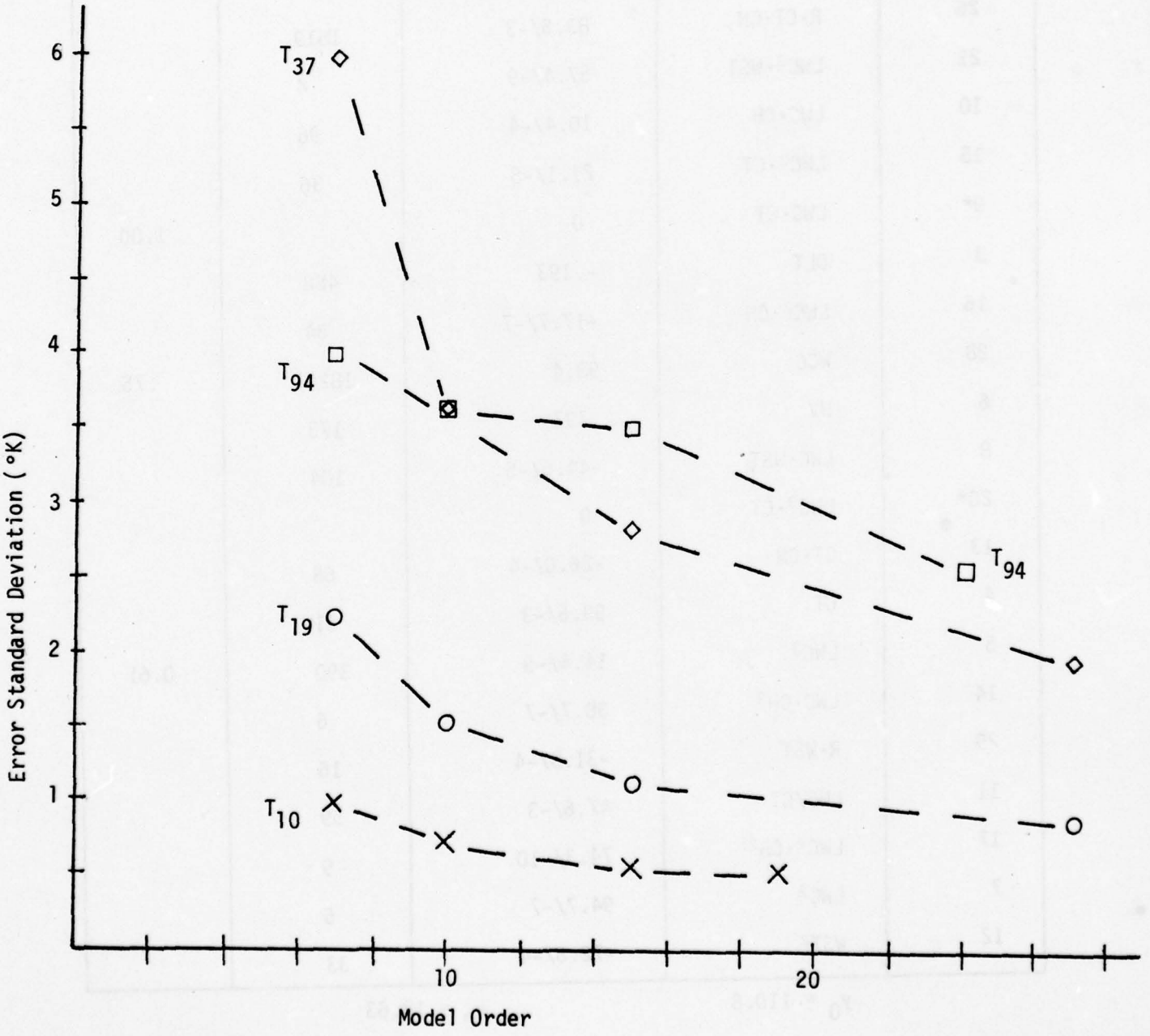


Table 3.4.4 Forward Model for T_{10}

#	Predictor	Coefficient	F	σ
1	LWC	34.3/-3	196	
31*	SCC	0		
26	R·CT·CH	83.5/-3	1519	
21	LWC^2·WST	57.4/-9	4	
10	LWC·CH	10.4/-4	96	
15	LWC^2·CT	21.1/-8	36	
9*	LWC·CT	0		1.00
3	DLT	-.193	488	
16	LWC^2·CH	-17.7/-7	84	
28	WCC	98.4	1829	.75
6	WV	.793	173	
8	LWC·WST	-44.5/-5	104	
20*	LWC^3·CT	0		
13	CT·CH	-26.0/-4	68	
4	CT	99.6/-3	31	
5	LWC^3	14.4/-9	390	0.61
14	LWC·CH^2	38.7/-7	6	
29	R·WST	-31.8/-4	16	
11	LWC/CT	47.6/-3	59	
17	LWC^2·CH^2	74.3/-10	9	
7	LWC^2	94.7/-7	5	
12	WST^2	-72.6/-4	33	

$y_0 = 110.8$ $\sigma_0 = 18.63$

Table 3.4.5 Forward Model for T_{19}

#	Predictor	Coefficient	F	σ
1	LWC	.116	57	
18	SCC	-18.6	8	
7	LWC^2	-23.4/-5	79	
22	$LWC \cdot WST^2$	-27.5/-5	11	
6	WV	9.63	554	
10	LWC·CH	20.2/-4	155	
9*	LWC·CT	0		2.25
27	R/CT	21.6	176	
20	$LWC^2 \cdot CT$	21.1/-10	272	
3	DLT	-23.2/-2	270	1.56
16	$LWC^2 \cdot CH$	-29.9/-7	456	
26	R·CT·CH	86.6/-3	302	
21	$LWC^2 \cdot WST$	76.9/-8	258	
17	WCC	314	30	
13	CT·CH	-72.7/-4	190	1.15
24	CT/WV	.146	8	
14	$LWC \cdot CH^2$	17.2/-6	41	
23	$LWC^2 \cdot WV$	10.3/-5	81	
29	R·WST	-22.0/-3	114	
28	R/WST	-3.26	98	.94

$y_0 = 115.9$ $\sigma_0 = 31.0$

Table 3.4.6 Forward Model for T_{37}

#	Predictor	Coefficient	F	σ
11	LWC/CT	-.148	11	
4	CT	.445	27	
19	SCC	-39.7	8	
5	LWC^2	29.4/-8	367	
10	LWC·CH	81.5/-4	388	
16	LWC^2·CH	-25.9/-6	419	
20*	LWC^2·CT	0		6.02
13	CT·CH	-13.5/-3	139	
17	$LWC^2 \cdot CH^2$	21.1/-8	227	
22	$LWC \cdot WST^2$	-90.5/-5	26	
8	LWC·WST	21.1/-3	18	3.66
7	LWC^2	-14.7/-4	513	
21	LWC^2·WST	20.2/-7	384	
6	WV	11.4	142	
3	DLT	-.327	114	
29*	R·WST	0		
23	LWC^2·WV	49.0/-5	218	2.87
25	$LWC \cdot WV^2$	-29.9/-3	149	
1	LWC	.412	150	

$y_0 = 134.8$ $\sigma_0 = 30.0$

Table 3.4.7 Forward Model for T_{94}

#	Predictor	Coefficient	F	σ
28	1/(1+LWC)	-662	1738	
29	$1/(1+LWC)^2$	2488	725	
5	CH	-.304	234	
12	1/CT	-19.0	594	
9	$CT \cdot CH^2$	-66.0/-5	183	
30	$WV^2/(1+LWC)$	3.55	119	
3	CT	.626	254	4.01
31	WCC	65.1	82	
32	$LWC \cdot WST^2$	-16.1/-4	51	
18	1/CH	-140.4	372	3.64
15*	LWC/CT^2	0		
7	WV	3.76	52	
25*	R/WST	-3.03	36	
13	CT·CH	49.1/-3	77	
14	$LWC^2 \cdot WST$	10.5/-7	86	3.52
27*	LWC^2/WV	0		
21	LWC/CH	2.24	516	
1	LWC	-.239	34	
4	CT	-2.06	223	
6	CT^2	38.5/-3	102	
8	LWC·WST	42.9/-3	46	

$y_0 = 278.2$ $\sigma_0 = 13.5$

3.4.3 Inversion Results

3.4.3.1 Kalman Filtering

Several studies have been carried out to evaluate the use of microwave data for estimating cloud, atmospheric and surface parameters. The effect of sequentially processing the data rather than batch processing in the extended Kalman Filters was studied. Since the lower frequency channel brightness temperatures are more linear in the data than the higher frequencies the sequential processing order was chosen as: T_{19}, T_{37}, T_{94}. The sequential processor is compared with the batch processor in Table 3.4.8. Notice that the IEKF sequential processor generally does a little worse. This suggests that performance improvements may be rather small and that the predominant remaining error sources may be due to modeling effects. The reduction of error with sequential processing for the EKF estimates of LWC and CH can be attributed directly to the fact that the most linear measurements are processed first.

A more detailed breakdown of the sequential processor performance is given in Table 3.4.9. Results are shown for errors after processing T_{19}, after processing T_{19} and T_{37}, and after processing T_{19}, T_{37} and T_{94} in sequence. Also shown in the table are the Cramer-Rao bounds on performance. The "information rms" is the upper bound on the information available about the variable of interest from the measurements. For a scalar problem, this quantity is the inverse of the lower bound on the rms error. The quantity "Cramer-Rao rms" is the lower bound on the achievable rms error for any unbiased estimator. It thus represents the best possible accuracy with which a variable may be estimated from the available data. Note that the error for DLT is near its bound and both are only slightly less than the

Table 3.4.8 INVERSION ERRORS USING KALMAN FILTER

- rms measurement noise in model = 1°
- noise-free data (803 meas)

			LWC	WST	DLT	CT	CH	WV
A priori		mean	64.6	13.64	-.25	6.46	32.0	1.73
		rms	46.9	5.70	1.92	5.76	14.0	.38
Simultaneous Updating	EKF	mean	38.8	-5.04	1.95	-.61	-23.0	.10
		rms	44.9	3.39	2.64	5.24	24.3	.39
	IEKF	mean	-4.04	-1.74	-.19	-.14	2.14	.01
		rms	13.77	2.62	1.79	5.12	10.62	.34
Sequential Updating	EKF	mean	-4.57	-2.73	-.47	-2.08	4.14	-.22
		rms	38.6	4.74	1.84	5.67	12.9	.42
T_{19} T_{37} T_{94}	IEKF	mean	.21	-1.20	.03	-.39	-3.37	-.15
		rms	16.4	2.92	1.85	5.09	14.1	.38

Table 3.4.9 INVERSION ERRORS FOR SEQUENTIAL PROCESSING

sequence: T_{19}, T_{37}, T_{94} (803 meas)

			LWC	WST	DLT	CT	CH	WV
		a priori mean	64.56	13.64	-.249	6.456	32.01	1.729
		a priori rms	46.92	5.70	1.918	5.759	13.99	.385
Iterated Extended Kalman Filter	processing T_{19} σ_r=1.	information rms	.189	2.364	.643	.248	.208	6.726
		Cramer-Rao rms	12.63	.512	1.737	4.862	7.686	.357
		error mean	-21.61	-.483	.020	-1.193	-1.752	-.009
		error rms	42.99	4.765	1.914	5.516	13.352	.378
	processing T_{37} σ_r=1.	information rms	2.537	3.059	.789	.346	.381	7.442
		Cramer-Rao rms	.409	.400	1.580	3.565	4.250	.198
		error mean	5.70	-2.200	.317	.170	-4.613	-.069
		error rms	16.43	2.383	1.972	4.981	14.260	.380
	processing T_{94} σ_r=1.	information rms	6.385	3.074	1.079	1.313	.597	8.522
		Cramer-Rao rms	.166	.395	1.144	1.270	2.813	.152
		error mean	.210	-1.202	.033	-.394	-3.374	-.151
		error rms	16.38	2.918	1.853	5.091	14.106	.377

a priori value. Thus, DLT is not, nor can it be, estimated with any confidence from the data.

It is useful from a design standpoint to investigate the Cramer-Rao bound in more detail. This is done in Table 3.4.10. Shown in this table are the bounds as a function of assumed rms measurement noise and modeling errors (σ_r); values of 0.5°, 2° and 5° were used, assuming equal noise rms values for each channel. The bound is dependent on the types of measurements used, but is independent of whether they are processed in batch or sequentially, or in which order they are processed. Several points should be made based on this data:

(1) T_{94} appears to be the best predictor for LWC, by a wide margin. The other measurements do not reduce the bound significantly. Note, however, that if the sensor errors are lower for low frequency channels, then T_{37} may be better than T_{94} for estimating LWC.

(2) Assuming that σ_r represents the effects of modeling errors, the importance of obtaining accurate forward models is clearly shown. LWC error for σ_r = 5° is nine times that for σ_r = 1°. WST error bounds are also increased dramatically.

(3) The bounds allow the designer to estimate the importance of a particular measurement directly. For example, it is clear that one cannot expect to achieve any significant performance improvements over the a priori errors for DLT or WV, assuming a 2° rms modeling error. In addition, if the rms modeling error is .5°, only marginal gains in estimating CH would be possible. However, it appears that large improvements relative to a priori rms errors for LWC and WST estimates are possible.

3.4.3.2 Nonlinear Regression

A series of inversions has been made on the new data base of 941 cases. The results are summarized in Table 3.4.11, which shows the rms inversion errors

Table 3.4.10 CRAMER-RAO BOUNDS FOR INVERSION ERRORS

• rms errors (863 meas)

		LWC	WST	DLT	CT	CH	WV
	a priori rms	99.3	5.72	1.85	7.29	14.2	.39
σ_r = 0.5°	T_{19}	9.5	.26	1.70	4.84	5.53	.31
	T_{37}	.22	.32	1.54	2.47	3.72	.31
	T_{94}	.10	3.70	1.47	.81	2.06	.22
	T_{19},T_{37}	.21	.20	1.46	2.37	2.72	.12
	T_{19},T_{94}	.10	.26	.80	.77	1.82	.09
	T_{37},T_{94}	.08	.31	.78	.71	1.64	.13
	T_{19},T_{37},T_{94}	.08	.20	.72	.67	1.50	.08
σ_r = 2°	T_{19}	17.3	1.00	1.79	4.90	10.40	.37
	T_{37}	.82	1.18	1.66	4.46	7.28	.37
	T_{94}	.37	4.96	1.61	2.57	6.52	.33
	T_{19},T_{37}	.81	.76	1.64	4.42	6.56	.29
	T_{19},T_{94}	.37	.94	1.51	2.48	6.00	.25
	T_{37},T_{94}	.33	1.13	1.48	2.28	5.16	.30
	T_{19},T_{37},T_{94}	.33	.75	1.45	2.23	4.92	.24
σ_r = 5°	T_{19}	31.2	2.29	1.84	5.10	12.43	.38
	T_{37}	2.00	2.49	1.76	4.87	10.84	.38
	T_{94}	.89	5.51	1.71	3.93	10.33	.36
	T_{19},T_{37}	1.99	1.69	1.75	4.85	10.33	.36
	T_{19},T_{94}	.88	2.06	1.68	3.89	9.83	.34
	T_{37},T_{94}	.81	2.43	1.66	3.77	8.98	.36
	T_{19},T_{37},T_{94}	.80	1.68	1.65	3.73	8.70	.34

Table 3.4.11 REGRESSION INVERSION ERRORS

std dev in °K (941 meas)

	LWC $\frac{g}{cm^2} \times 10^3$	WST $\frac{m}{s}$	DLT °K	CT $m \times 10^2$	CH $m \times 10^2$	WV $\frac{g}{cm^2}$	R $\frac{mm}{hr}$
a priori	312.2	5.75	2.21	9.80	13.55	0.477	5.78
linear	111 3/4	4.33 3/4	2.17 4/4	5.35 3/4	10.88 4/4	.27 4/4	2.54 4/4
nonlinear	47.34 6/30 45.52 8/30 43.11 10/30	1.75 7/30 1.58 10/30 1.35 15/30	2.12 4/30	4.9 5/30 4.75 8/30	8.1 10/30 7.4 16/30	.17 7/30 .124 13/30	1.98 6/30

for both linear and nonlinear regressions. The notation "a/b" in the table indicates that "a" out of a total of "b" possible predictors were selected by the stepwise regression algorithm. A listing of the total set of 30 predictors tried for inversion is given in Table 3.4.12, along with the mean and rms value of each predictor.

The individual regression models are given in Tables 3.4.13 - 3.4.19, along with F ratio and the partial correlations before entering the first predictor. Also shown are the rms fit errors at the points corresponding to values in Table 3.4.11.

Table 3.4.12(a) REGRESSION PREDICTOR STATISTICS 941 Cases

Variable	Mean	σ	#
T_{10}	129.1	18.63	1
T_{19}	164.2	31.02	2
T_{37}	202.8	29.97	3
T_{94}	259.2	13.46	4
$1/T_{10}$	.00787	0.00090	5
$1/T_{19}$	.00626	0.00091	6
$1/T_{37}$	.00503	0.00069	7
$1/T_{94}$	.00387	0.00022	8
$T_{10} \cdot T_{19}$	21753	8248	9
$T_{10} \cdot T_{37}$	26646	8100	10
$T_{10} \cdot T_{94}$	33525	5539	11
$T_{19} \cdot T_{37}$	34139	12272	12
$T_{19} \cdot T_{94}$	42693	9035	13
$T_{37} \cdot T_{94}$	52794	9414	14
T_{10}/T_{19}	0.792	0.0375	15
T_{10}/T_{37}	0.640	0.0500	16
T_{10}/T_{94}	0.498	0.0689	17
T_{19}/T_{10}	1.266	0.0636	18
T_{19}/T_{37}	1.573	0.1226	19
T_{19}/T_{94}	2.038	0.2285	20

Table 3.4.12(b)

Variable	Mean	σ	#
T_{37}/T_{10}	0.808	0.0553	21
T_{37}/T_{19}	0.633	0.1110	22
T_{37}/T_{94}	1.243	0.0794	23
T_{94}/T_{10}	1.617	0.2164	24
T_{94}/T_{19}	0.781	0.0979	25
T_{94}/T_{37}	1.298	0.1436	26
$T_{10}\cdot T_{10}$	17009	5607	27
$T_{19}\cdot T_{19}$	27909	12169	28
$T_{37}\cdot T_{37}$	42022	12942	29
$T_{94}\cdot T_{94}$	67378	6714	30
LWC	165.6	312.2	
WST	13.69	5.75	
DLT	0	2.21	
CT	9.44	9.80	
CH	32.57	13.55	
WV	1.845	0.477	
R	1.701	5.775	

Table 3.4.13 INVERSION FOR LWC

Predictor	Coeff	F	corr	σ
$T_{10} \cdot T_{19}$	0.546	385	0.949	
$1/T_{10}$	0		-0.844	
$T_{19} \cdot T_{19}$	-0.470	227	0.945	
$T_{37} \cdot T_{37}$	0		0.805	
$T_{10} \cdot T_{10}$	0		0.940	
T_{19}/T_{94}	0		-0.766	47.3
$T_{10} \cdot T_{37}$	-0.434	149	0.910	
T_{10}/T_{19}	0		-0.652	45.5
T_{10}/T_{94}	-5678	16	0.854	
$T_{19} \cdot T_{37}$	0.471	219	0.903	
T_{37}/T_{94}	7202	135	-0.592	
T_{37}	-51.7	41	0.769	
T_{94}/T_{19}	-3211	7	0.805	
T_{37}/T_{19}	11.8/3	37	0.907	
$T_{10} \cdot T_{94}$	71.5/-3	8	0.879	43.1

$y_o = -6107$ $\sigma_o = 312.2$

Table 3.4.14 INVERSION FOR WST

Predictor	Coeff	F	corr	σ
$1/T_{10}$	-68.0/4	304	-0.495	
$T_{19} \cdot T_{19}$	83.7/-3	150	0.235	
T_{37}/T_{94}	-1436	104	-0.282	
$1/T_{94}$	-12.0/4	37	-0.103	
T_{10}/T_{94}	729	29	0.366	1.75
T_{10}/T_{19}	-0.16	165	0.282	
T_{10}	1.94	< 1	0.388	
$T_{10} \cdot T_{37}$	0.12	269	0.296	1.28
T_{37}	-16.0	223	0.247	
$T_{10} \cdot T_{10}$	62.7/-3	220	0.334	
$T_{19} \cdot T_{37}$	-37.5/-3	142	0.240	
T_{19}	-18.9	61	0.288	
$T_{19} \cdot T_{37}$	1810	193	-0.239	
$1/T_{19}$	55.5/4	170	-0.395	
T_{94}/T_{10}	-35.7	< 1	-0.385	
T_{94}	0.748	13	0.098	1.35
T_{10}/T_{19}	-6073	27	-0.003	
T_{94}/T_{37}	-179	27	-0.306	1.25

y_o = 10094 σ_o = 5.75

Table 3.4.15 INVERSION FOR DLT

Predictor	Coeff	F	corr	σ
$T_{94} \cdot T_{94}$	0		0.109	
T_{94}	0.32	49	0.103	
$1/T_{37}$	12.8/3	39	0.030	
T_{94}/T_{10}	-3.35	13	0.061	
T_{94}/T_{37}	-41.9	28	0.082	2.12

y_o = -87.5 σ_o = 2.21

Table 3.4.16 INVERSION FOR CT

Predictor	Coeff	F	corr	σ
$1/T_{10}$	9140	29	-0.742	
T_{94}/T_{37}	115.9	102	-0.734	
$T_{10} \cdot T_{19}$	15.3/-4	22	0.814	
$T_{19} \cdot T_{37}$	-48.3/-4	21	0.819	
$T_{37} \cdot T_{37}$	38.8/-4	63	0.766	
T_{19}	2.92	96	0.819	
$T_{19} \cdot T_{94}$	-70.2/-4	35	0.779	
$T_{94} \cdot T_{94}$	87.6/-5	7	0.169	4.75

$y_o = -482.5$ $\sigma_o = 9.80$

Table 3.4.17 INVERSION FOR CH

Predictor	Coeff	F	corr	σ
T_{94}/T_{37}	-1740	64	-0.397	
T_{37}/T_{37}	0.147	119	0.232	
T_{10}/T_{37}	-10.6/3	202	-0.07	
$T_{94} \cdot T_{94}$	16.8/-3	15	-0.242	
$T_{10} \cdot T_{94}$	51.7/-3	93	0.114	
T_{94}/T_{19}	-8868	145	0.369	
$T_{37} \cdot T_{94}$	-0.121	99	0.128	
$T_{19} \cdot T_{19}$	40.6/-3	20	0.198	
T_{37}/T_{94}	-4751	95	-0.014	8.09
$T_{19} \cdot T_{37}$	-0.16	97	0.217	
T_{10}/T_{19}	7911	173	-0.159	
$T_{19} \cdot T_{94}$	41.5/-3	96	0.131	
T_{37}/T_{19}	6266	56	0.285	
T_{19}/T_{37}	2276	70	0.086	
T_{19}/T_{10}	-1716	14	0.158	
T_{94}	-2.39	4	-0.229	7.40

$y_0 = 10,823$ $\sigma_0 = 13.55$

Table 3.4.18 INVERSION FOR WV

Predictor	Coeff	F	corr	σ
T_{19}/T_{10}	0		0.702	
T_{37}/T_{94}	-73.3	430	-0.485	
T_{94}/T_{19}	-89.0	394	0.567	
T_{19}/T_{37}	57.2	360	0.028	
T_{10}/T_{19}	-403	754	-0.701	0.170
$1/T_{19}$	17.6/3	123	-0.655	
$1/T_{10}$	-15.7/3	107	-0.584	
T_{37}/T_{10}	-75.5	281	0.509	
T_{94}/T_{10}	73.3	102	-0.614	
T_{19}/T_{94}	-64.9	112	-0.479	
T_{10}/T_{94}	246	323	0.527	
T_{37}/T_{19}	-129	412	0.650	
$T_{19} \cdot T_{94}$	-31.5/-5	39	0.681	
$T_{10} \cdot T_{10}$	30.9/-5	20	0.621	0.124

$y_o = 446.3$ $\sigma_o = 0.477$

Table 3.4.19 INVERSION FOR R

Predictor	Coeff	F	corr	σ
$T_{19}\ T_{19}$	14.6/-3	126	0.868	
T_{19}	-47.1/-1	201	0.834	
$T_{10}\ T_{19}$	-16.4/-3	27	0.858	
$1/T_{19}$*			-0.735	
T_{10}/T_{19}	-42.2/1	137	-0.593	
$T_{10}\ T_{10}$	11.6/-3	25	0.836	
$1/T_{10}$	-37.2/3	155	-0.750	1.98

y_o = 1154 σ_o = 5.78

3.5 Inversion Errors for Data Set #3

3.5.1 Data Base

Data Base #2 was modified in several ways to produce Data Base #3. The number of thin clouds with high LWC was reduced to more closely match their small probability of occurrence. The frequency of high rain rates (> 12 mm/hr) was reduced and several low rain rate cases (< 3 mm/hr) were added. Rain layers up to 4000 meters thick were added. The clear column cases were removed so that all cases were run over clouds. Since ice clouds cannot be detected from microwave measurements, these cases were eliminated from the data set, leaving only water clouds. The summary of cases for Data Set #3 is given in Table 3.5.1. The resulting parameter correlation matrix is given in Table 3.5.2.

3.5.2 Forward Models

The forward models determined via nonlinear regression are given in Table 3.5.3 for all four microwave channels and are seen to be highly nonlinear. The significance of the predictors can be determined by examining Table 3.5.4, which indicates the order in which the predictors were entered into the models and gives the F ratios for the final models.

TABLE 3.5.1

Summary of Data Set #3

for Microwave Measurements

Wind Speed (m/sec)

	8	10	12.50	15	17.50	20	22.5
#	311	20	10	20	10	281	10

Cloud Thickness (m)

	330	500	660	1000-2350	2500	2600-3600	3700-4200
#	20	322	20	60	132	54	54

Cloud Top Height (m)

	660	1320	1500	2500	2900	3000	3400	4000	4500	4900
#	20	20	56	14	190	40	2	66	32	222

Water Vapor (gm/m^2)

	.89-1.02	1.38-1.62	1.70-1.93	2.01-2.48	2.61-2.89
#	40	100	280	102	140

Rain Rate (mm/hr)

	0	1	2.4	3	5	7	12	15	24	30
#	498	12	24	12	16	22	24	16	22	16

TABLE 3.5.2

Parameter Correlation Matrix

for Data Set #3

	LWC	WST	DLT	CT	CH	WV	R
LWC	427.0						
WST	0.020	5.83					
DLT	0	0	2.09				
CT	0.845	0.028	0	12.20			
CH	0.245	0.041	0	0.436	12.85		
WV	0.735	0.014	~ 0	0.695	0.031	0.539	
R	0.766	0.014	0	0.615	0.096	0.619	6.80
mean	291.4	13.80	0	15.09	35.01	2.012	2.83

Note: diagonal elements are rms values

Table 3.5.3

MICROWAVE FORWARD REGRESSION MODELS

•662 meas

Term	T_{10}		T_{19}		T_{37}		T_{94}	
const.	108.6		120.5		149.4		279.0	
1	96.08E-3	LWC	.2040	LWC	.3836	LWC	-41.72E-3	LWC
2	2.054	WV	7.332	WV	-.5965	CT	.4996	DLT
3	-12.36E-3	LWC•CT	-81.73E-6	LWC^2	28.74E-8	LWC^3	-.5085	CT
4	10.01E-3	LWC•CH	-20.92E-4	LWC•WST	3.396	WV	20.93E-3	CT^2
5	-.6145	LWC/CT	-38.91E-4	LWC•CT	-62.10E-5	LWC^2	-17.79E-5	$CT•CH^2$
6	-11.64E-5	$LWC•CH^2$	27.02E-4	LWC•CH	.9861	LWC/CT	-59.57	1/CT
7	16.32E-5	$LWC•CT^2$	-17.56E-2	LWC/CT	22.72E-3	CT•CH	63.62E-8	LWC^2•WST
8	-33.15E-7	LWC^2•WV	34.44E-7	LWC^2•CT	49.01E-7	LWC^2•CT	2.855	LWC/CT^2
9	.1134	R•(CH-CT)	-24.97E-7	LWC^2•CH	-33.55E-9	$LWC^2•CH^2$	-31.68	1/CH
10	-18.73E-3	R•WST	22.76	SCC	21.35	SCC	1.091	LWC/CH
11	16.03	SCC	75.45E-8	$LWC^2•WST^2$	-30.89E-10	LWC^3•CT	-516.0	1/LWC
12			.1209	R•(CH-CT)	17.54E-7	LWC^2•WST	1449	$1/LWC^2$
13					-12.69E-5	$LWC•WST^2$	22.36	WV^2/LWC
14					34.99E-3	R•(CH-CT)	-41.17E-6	$LWC•WST^2$
15							5.367	SCC

TABLE 3.5.4

Stepwise Regression Predictors

for Microwave Forward Models

• Data Set #3

• F ratios are for final model

Step #	T_{10}		T_{19}		T_{37}		T_{94}	
	pred	F	pred	F	pred	F	pred	F
1	LWC	71	LWC	719	CT	55	1/LWC	1075
2	SCC	3167	LWC^2	215	LWC/CT	301	$1/LWC^2$	521
3	R(CH-CT)	320	SCC	3791	$LWC^2 \cdot WV$		$CT \cdot CH^2$	111
4	R•WST	66	LWC•WST	185	SCC	3576	1/CT	68
5	$LWC^2 \cdot WV$	84	R(CH-CT)	577	$LWC \cdot WST^2$	573	SCC	469
6	LWC•CH	334	LWC^3		$LWC^2 \cdot CT$	418	WV^2/LWC	185
7	LWC/CT	69	a		LWC	3650	DLT	149
8	WV	33	LWC•CH	469	$LWC^3 \cdot CT$	298	$LWC \cdot WST^2$	285
9	$LWC \cdot CH^2$	269	$LWC^2 \cdot CH$	184	LWC^2	1970	CT^2	108
10	LWC•CT	383	WV	383	$-LWC^2 \cdot WV$		$LWC^2 \cdot WST$	193
11	$LWC \cdot CT^2$	354	-a		LWC^3	1026	LWC	351
12			$LWC^2 \cdot WST$	45	$LWC^2 \cdot WST$	260	LWC/CH	180
13			LWC/CT	10	$LWC^2 \cdot CH^2$	526	LWC/CT^2	69
14			$LWC^3 \cdot CT$		CT•CH	319	1/CH	36
15			LWC•CT	486	WV	85	CT	16
16			$-LWC^3$		R(CH-CT)	48		
17			$LWC^2 \cdot CT$	328				
18			$-LWC^3 \cdot CT$					

a = 1/LWC + 1000 WV

3.5.3 Inversion Results

Inversions were carried out using nonlinear regressions with a single channel and noise-free data. The results are shown in Table 3.5.5. Predictors tried included powers, inverses, exponentials and logarithms of the basic predictor. Terms employed in the model, at the 5% significance level, are indicated by x. The resulting rms errors are shown at the bottom, along with the apriori values. The values 1, 2, 3, 4 denote the 10.7, 19, 37 and 94 GHz channels, respectively.

Next, a series of linear regressions using all four channels were run using noise-free data on both fitting and testing sets. The results are shown in Tables 3.5.6 - 3.5.12. It can be seen that R^2 for LWC and CT is reasonably high, WV and R values are moderate and CH and WST values are low; R^2 for DLT is less than 2%, indicating, as expected, that atmospheric temperature variations cannot be predicted accurately from microwave measurements over water clouds.

A series of nonlinear regressions were run to assess the possibility of improving the regression inversions. The results are shown in Tables 3.5.13 - 3.5.19. In each table, the predictors are listed in order of entry into the model, together with the regression coefficient at the final step. The running values of goodness-of-fit (R^2) in percent and rms fit error are also given. An asterisk denotes a predictor that was subsequently deleted from the model at the 5% significance level.

One of the problems encountered in regression analysis is non-robustness in the presence of noise. High order regression models have a tendency to overfit the data, resulting in excessively large errors in

the presence of noise in the test set (i.e., in practice). An indication of this problem is given in Tables 3.5.20 - 3.5.26 which show the effect of adding noise to the test set with rms values of 1° and 5° over all channels. The noise on each channel was independent of the noise on all other channels. The data show the actual rms errors for the test set for increasing-order models. Note that for low-order (less than 3 or 4) models, there is generally only a small degradation of performance. However, degradation can become severe at high orders (generally greater than 4 or 5). A comparison of noise sensitivity for both regression and Kalman Filtering methods is given in Tables 3.5.27 - 3.5.32. Two values of rms measurement noise were used for the Kalman Filters: σ_r = 0.5°, 2°; i.e., $V = (0.5)^2 I_4$ or $(2)^2 I_4$, with I_4 the 4x4 identity matrix. It can be seen that the nonlinear regression method is much more sensitive to noise than the Kalman Filters, even for 1° rms noise. Note also that the LWC errors for the Kalman Filters are minimized by assuming an rms measurement error of about 4°, which accounts principally for modeling error.

Another important aspect of inversion is the possiblity of data dropout, which could be caused by loss of communication, loss of significance (bits), intermittent sensor problems, etc. In this case, something unexpected has happened; however, it is still desirable to be able to process the remaining data. In the regression approach, the inversion depends on all of the data being present. For absent data, one could simply use an apriori mean. Of course, one can also produce regression models for all possible combinations; however, this would be an inefficient procedure in practice. The Kalman Filter method takes missing data into

account in a very natural way. The effect of missing data (one channel out) on regression errors is shown in Tables 3.5.33 - 3.5.38 for both fitting to the missing data set (model) and using the apriori mean (ave). The (ave) values for regression, denoted by "reg", are compared to Kalman Filter regression in Tables 3.5.39 - 3.5.40. Both mean and rms errors are given and show that the Iterated Extended Kalman Filter is the most tolerant to missing data.

3.5.4 Cramer-Rao Bounds

The preceding results have been predicated on several specific inversion techniques. However, none of these can be said to be truly optimum. For example, there are many existing Bayesian techniques which could not be studied, due to the limited scope of this study. Rather than search for the "best" overall method, it is helpful to first consider the question of achievable performance, given the available sources of information. That is, we wish to consider the extent to which the apriori errors can be reduced using the four selected microwave channels. A technique which can be used for this purpose is the Cramer-Rao bound. As shown in Appendix B, lower bounds on the rms estimation errors for each parameter to be inverted for can be computed from knowledge of the measurement function $h(\cdot)$ and the statistics of the measurement errors (v). A summary of the achievable rms errors is given in Tables 3.5.41 and 3.5.42 for all sensor combinations. The sensor rms errors were σ_r = 0.5°, 1.0°, 2.0°, and 5.0°. In each case the errors for each channel were independent of the other channel errors, but the rms values were identical for all channels. It should be noted that the bounds are independent of the order in which the data are processed.

TABLE 3.5.5

SUMMARY OF INVERSIONS USING

SINGLE MICROWAVE CHANNELS

• 662 meas. over water clouds

	LWC				WST				DLT				CT				CH				WV				R			
NT	1	2	3	4	1	2	3	4	1	2	3	4	1	2	3	4	1	2	3	4	1	2	3	4	1	2	3	4
T	X	X		X												X					X		X			X	X	
1./T			X		X		X						X															
T^2	X		X	X											X							X	X		X	X	X	
$1./T^2$					X	X							X	X	X													
T^3	X	X	X	X	X	X									X								X	X	X		X	X
$1./T^3$				X	X	X	X	X					X	X	X		X	X	X				X	X	X	X		X
e^T		X	X	X							X	X		X		X			X	X	X		X	X		X	X	X
e^{-T}					X	X							X					X				X	X	X				
Log T														X									X				X	
1./Log T																												
error r.m.s.	107.05	106.79	196.95	377.84	4.42	4.79	5.27	5.77			2.07	2.06	6.50	5.51	5.50	10.92	12.14	12.03	11.81	12.76	.38	.35	.36	.46	4.18	3.66	4.29	6.37
a priori r.m.s.	427.0				5.83				2.09				12.20				12.85				0.54				6.98			

Table 3.5.6

INVERSION FOR INTEGRATED LIQUID WATER VIA LINEAR REGRESSION USING MICROWAVE MEASUREMENTS

Variable in Order of Entry	Coefficient (Final)	R^2 (At Step)	σ (At Step)
mean	2.911×10^{2} [a]	0	427.0
T_{19}	1.723×10^{1}	.861	159.1
T_{37}	-7.312	.898	136.5
T_{94}	1.505	.898	136.1
T_{10}	-2.223	.899	135.6

Intercept: -1.298×10^{3}

Table 3.5.7

INVERSION FOR WIND SPEED VIA LINEAR REGRESSION USING MICROWAVE MEASUREMENTS

Variable in Order of Entry	Coefficient (Final)	R^2 (At Step)	σ (At Step)
mean	1.380×10^{1} [a]	0	5.83
T_{10}	4.965×10^{-1}	7.8	5.59
T_{19}	-3.820×10^{-1}	16.1	5.34
T_{37}	1.298×10^{-1}	21.3	5.17

Intercept: -1.388×10^{1}

[a] mean value

Table 3.5.8

INVERSION FOR DLT VIA LINEAR
REGRESSION USING MICROWAVE MEASUREMENTS
662 meas. over water clouds

Variable in Order of Entry	Coefficient (Final)	R^2 (At Step)	σ (At Step)
mean	0[a]	0	2.09
T_{94}	3.123×10^{-2}	.009	2.09
T_{37}	-5.638×10^{-3}	.016	2.08

Intercept: -6.847

[a]mean value

Table 3.5.9

INVERSION FOR CLOUD THICKNESS VIA LINEAR
REGRESSION USING MICROWAVE MEASUREMENTS
662 meas. over water clouds

Variable in Order of Entry	Coefficient (Final)	R^2 (At Step)	σ (At Step)
mean	1.509×10^{1} [a]	0	12.20
T_{37}	1.638×10^{-1}	.770	5.85
T_{94}	-1.623×10^{-1}	.794	5.53
T_{19}	2.122×10^{-1}	.803	5.41
T_{10}	-1.393×10^{-1}	.807	5.36

Intercept: 3.484

[a]mean value

Table 3.5.10

INVERSION FOR CLOUD TOP HEIGHT VIA LINEAR REGRESSION USING MICROWAVE MEASUREMENTS

662 meas. over water clouds

Variable in Order of Entry	Coefficient (Final)	R^2 (At Step)	σ (At Step)
mean	3.501×10^{1} [a]	0	12.85
T_{37}	5.908×10^{-1}	.131	11.98
T_{94}	-6.042×10^{-1}	.217	11.38
T_{19}	-6.492×10^{-1}	.265	11.02
T_{10}	4.939×10^{-1}	.299	10.76

Intercept: 1.128×10^{2}

[a] mean value

Table 3.5.11

INVERSION FOR ATMOSPHERIC WATER VAPOR VIA LINEAR REGRESSION USING MICROWAVE MEASUREMENTS

662 meas. over water clouds

Variable in Order of Entry	Coefficient (Final)	R^2 (At Step)	σ (At Step)
mean	2.01×10^{-1}	0	0.539
T_{19}	3.816×10^{-2}	.575	0.352
T_{10}	-2.633×10^{-2}	.589	0.346
T_{37}	-1.618×10^{-2}	.627	0.329
T_{94}	1.343×10^{-2}	.665	0.312

Intercept: -1.183

[a] mean value

Table 3.5.12

INVERSION FOR RAIN RATE VIA LINEAR

REGRESSION USING MICROWAVE MEASUREMENTS

662 meas. over water clouds

Variable in Order of Entry	Coefficient (Final)	R^2 (At Step)	σ (At Step)
mean	2.83[a]	0	6.98
T_{19}	5.223×10^{-1}	54.7	4.58
T_{37}	-2.846×10^{-1}	64.1	4.08
T_{10}	-2.725×10^{-1}	68.5	3.89
T_{94}	1.196×10^{-1}	69.2	3.77

Intercept: -2.298×10^{1}

[a]mean value

Table 3.5.13

REGRESSION FOR LWC

Variable in Order of Entry	Coefficient (Final)	R^2 (At Step)	σ (At Step)
CONST	2.914×10^{2} [a]	0	427.0
T_{19}^2	-4.163×10^{-2}	.896	138.1
$1/T_{19}$	6.063×10^{5}	.929	114.1
$\exp(T_{19})$	-3.953×10^{-117}	.937	107.7
$T_{10} \cdot T_{37}$ [b]	0	.945	100.5
$\exp(T_{37})$	-1.528×10^{-118}	.948	97.5
T_{19}	3.821×10^{1}	.950	96.0
T_{94}/T_{10}	1.694×10^{3}	.950	95.5
T_{10}/T_{94}	6.861×10^{3}	.955	91.4
$T_{37} \cdot T_{94}$	7.077×10^{-3}	.957	89.4

Intercept: -1.598×10^{4}

[a] mean value

[b] deleted in later step

Table 3.5.14

REGRESSION FOR WST

Variable in Order of Entry	Coefficient (Final)	R^2 (At Step)	σ (At Step)
CONST	1.380×10^{1} [a]	0	5.83
$1/T_{10}$	-7.497×10^{4}	.124	5.46
T_{19}^2	8.611×10^{-3}	.722	3.08
$T_{10} \cdot T_{37}$	8.985×10^{-3}	.754	2.90
$\exp(T_{37})$	6.642×10^{-120}	.769	2.81
$\exp(T_{19})$	-4.161×10^{-119}	.779	2.75
T_{37}/T_{19}	-4.204×10^{2}	.798	2.63
T_{10}/T_{94}	-2.166×10^{2}	.828	2.43
$1/T_{37}$ [b]	0	.832	2.40
T_{94}/T_{37}	-2.188×10^{2}	.841	2.34
T_{19}	-8.762	.860	2.19
T_{10}/T_{19}	-7.164×10^{2}	.868	2.13
T_{94}/T_{19}	1.155×10^{2}	.873	2.09

Intercept: 2.858×10^{3}

---

[a] mean value

[b] deleted in later step

Table 3.5.15

REGRESSION FOR DLT

Variable in Order of Entry	Coefficient (Final)	R^2 (At Step)	σ (At Step)
CONST	0[a]	0	2.09
$\exp(T_{94})$	2.380×10^{-123}	.034	2.06
$\exp(T_{37})$	3.663×10^{-120}	.058	2.04

Intercept: -7.821×10^{-2}

[a] mean value

Table 3.5.16

REGRESSION FOR CT

Variable in Order of Entry	Coefficient (Final)	R^2 (At Step)	σ (At Step)
CONST	1.509×10^{1} [a]	0	12.20
T_{37}/T_{94}	7.329×10^{1}	.795	5.53
T_{94}/T_{37} [b]	0	.808	5.35
$\exp(T_{19})$	-2.329×10^{-119}	.813	5.28
T_{10}^2	-1.127×10^{-2}	.821	5.17
$1/T_{10}$	7.866×10^{4}	.840	4.90
$1/T_{19}$	-3.293×10^{4}	.844	4.83
T_{19}/T_{10}	-1.230×10^{2}	.847	4.80
T_{10}	5.906	.850	4.76

Intercept: -8.737×10^{2}

---

[a] mean value

[b] deleted in later step

Table 3.5.17

REGRESSION FOR CH

Variable in Order of Entry	Coefficient (Final)	R^2 (At Step)	σ (At Step)
CONST	3.501×10^{1} [a]	0	12.85
T_{94}/T_{37} [b]	0	.189	11.59
T_{37}^2	4.940×10^{-2}	.346	10.41
T_{94}/T_{19}	-5.450×10^{2}	.377	10.17
T_{37}/T_{19}	-2.531×10^{3}	.500	9.11
T_{19}/T_{37}	-5.371×10^{3}	.541	8.74
T_{19}^2	1.495×10^{-2}	.613	8.03
T_{19}/T_{94}	-3.268×10^{1}	.630	7.86
$T_{19} \cdot T_{37}$	-6.058×10^{-2}	.638	7.78
T_{94}/T_{10}	2.492×10^{2}	.654	7.60
T_{19}/T_{10}	-2.572×10^{2}	.660	7.54
$T_{37} \cdot T_{94}$	-2.370×10^{5}	.669	7.45

Intercept: 8.423×10^{3}

---

[a] mean value

[b] deleted in later step

Table 3.5.18

REGRESSION FOR WV

Variable in Order of Entry	Coefficient (Final)	R^2 (At Step)	σ (At Step)
CONST	2.012[a]	0	.539
$T_{19} \cdot T_{94}$ [b]	0	.591	.345
$1/T_{10}$	4.837×10^{3}	.632	.328
T_{37}/T_{19}	-1.469×10^{1}	.705	.294
$1/T_{19}$	-3.900×10^{3}	.740	.276
T_{19}/T_{10}	-1.728×10^{1}	.824	.227
$T_{37} \cdot T_{94}$	3.214×10^{-4}	.834	.221
T_{94}/T_{37}	-6.536	.839	.217
T_{37}/T_{94}	-1.941×10^{1}	.852	.208
$\exp(T_{37})$	-3.600×10^{-121}	.856	.206
$T_{10} \cdot T_{94}$	-5.796×10^{-4}	.857	.205
T_{10}	1.520×10^{-1}	.861	.202

Intercept: 3.480×10^{1}

---

[a] mean value

[b] deleted in later step

Table 3.5.19

REGRESSION FOR R

Variable in Order of Entry	Coefficient (Final)	R^2 (At Step)	σ (At Step)
CONST	2.83[a]	0	6.80
T_{19}^2	5.777×10^{-3}	.590	4.36
$T_{10} \cdot T_{37}$	-5.828×10^{-4}	.703	3.71
$\exp(T_{37})$	1.058×10^{-119}	.730	3.54
T_{94}/T_{19}[b]	0	.752	3.39
T_{10}^2	6.223×10^{-3}	.756	3.37
T_{37}/T_{19}	2.754×10^{2}	.770	3.28
$1/T_{10}$	-8.053×10^{2}	.774	3.25
$T_{10} \cdot T_{19}$	-1.129×10^{-2}	.778	3.22
T_{19}/T_{37}	3.409×10^{2}	.781	3.20
T_{37}/T_{94}	-7.894×10^{1}	.785	3.17
T_{94}/T_{10}	-2.075×10^{1}	.789	3.14

Intercept: -5.120×10^{2}

---

[a] mean value

[b] deleted in later step

TABLE 3.5.20

EFFECT OF NOISE ON LWC REGRESSION

RMS Errors

$\sigma_{model} = 0$

Order	$\sigma_n = 0$	$\sigma_n = 1°$	$\sigma_n = 5°$
0	427.0	427.0	427.0
1	138.0	138.7	147.1
2	114.0	114.7	123.6
3	107.5	111.0	182.4
4	100.2	111.0	342.9
5	95.6	110.6	6649
6	95.1	109.7	6393
7	89.1	108.3	6405
8	85.8	119.2	6924
9	84.5	122.8	4453
10	82.7	140.4	3386

TABLE 3.5.21

EFFECT OF NOISE ON WST REGRESSION

RMS Errors

$\sigma_{model} = 0$

Order	$\sigma_n = 0$	$\sigma_n = 1°$	$\sigma_n = 5°$
0	5.83	5.83	5.83
1	5.45	5.45	5.46
2	3.07	3.31	6.57
3	2.89	3.11	6.34
4	2.80	3.11	198.9
5	2.74	3.12	203.8
6	2.62	3.29	265.9
7	2.42	3.20	232.7
8	2.38	3.22	222.3
9	2.32	3.13	186.4
10	2.11	2.86	141.8
11	2.07	2.77	149.8
12	2.04	2.76	149.1
13	2.02	2.91	132.5

TABLE 3.5.22

EFFECT OF NOISE ON DLT REGRESSION

RMS Errors

$\sigma_{model} = 0$

Order	$\sigma_n = 0$	$\sigma_n = 1°$	$\sigma_n = 5°$
0	2.09	2.09	2.09
1	2.06	2.54	5196
2	2.03	2.52	5246

TABLE 3.5.23

EFFECT OF NOISE ON CT REGRESSION

RMS Errors

$\sigma_{model} = 0$

Order	$\sigma_n = 0$	$\sigma_n = 1°$	$\sigma_n = 5°$
0	12.20	12.20	12.20
1	5.52	5.50	5.75
2	5.34	5.32	5.59
3	5.27	5.28	6.60
4	4.88	5.19	14.19
5	4.82	5.13	14.24
6	4.74	4.78	6.92
7	4.67	4.75	7.30

TABLE 3.5.24

EFFECT OF NOISE ON CH REGRESSION

RMS Errors

$\sigma_{model} = 0$

Order	$\sigma_n = 0$	$\sigma_n = 1°$	$\sigma_n = 5°$
0	12.85	12.85	12.85
1	11.58	11.59	11.67
2	10.39	10.43	11.59
3	10.14	10.22	11.95
4	9.09	9.45	15.67
5	8.71	9.19	17.48
6	7.99	8.62	18.28
7	7.75	8.58	20.11
8	7.56	8.40	20.84
9	7.40	8.31	21.09
10	7.31	8.13	21.77

TABLE 3.5.25

EFFECT OF NOISE ON WV REGRESSION

RMS Errors

$\sigma_{model} = 0$

Order	$\sigma_n = 0$	$\sigma_n = 1°$	$\sigma_n = 5°$
0	.539	.539	.539
1	.345	.345	.350
2	.327	.328	.357
3	.293	.298	.422
4	.275	.291	.574
5	.226	.283	.920
6	.220	.276	.936
7	.220	.276	.980
8	.216	.275	9.36
9	.207	.278	8.88
10	.204	.277	8.13

TABLE 3.5.26

EFFECT OF NOISE ON R REGRESSION

RMS Errors

$\sigma_{model} = 0$

Order	$\sigma_n = 0$	$\sigma_n = 1°$	$\sigma_n = 5°$
0	6.80	6.80	6.80
1	4.35	4.36	4.41
2	3.70	3.76	4.71
3	3.53	3.62	298
4	3.38	3.46	286
5	3.26	3.35	313
6	3.23	3.33	314
7	3.20	3.29	271
8	3.18	3.28	278
9	3.12	3.20	236

TABLE 3.5.27

INTEGRATED LIQUID WATER INVERSION

ERRORS USING MICROWAVE MEASUREMENTS

• 662 meas over water clouds

		Regression		Kalman Filter	
		linear	nonlinear	EKF	IEKF
a priori		427.0 (0)	427.0 (0)	427.0 (0)	427.0 (0)
$\sigma_r = 0.5°$	$\sigma_n = 0$	135.6	88.9	292.1 (231.1)	306.1 (89.6)[a]
	$\sigma_n = 1°$		140.4	291.3 (231.4)	316.5 (100.6)
	$\sigma_n = 5°$		3386	291.6 (232.9)	349.0 (113.8)
$\sigma_r = 2°$	$\sigma_n = 0$	135.6	88.9	287.3 (225.4)	235.1 (76.2)
	$\sigma_n = 1°$		140.4	286.6 (225.7)	228.7 (72.0)
	$\sigma_n = 5°$		3386	286.8 (227.1)	293.0 (92.0)
$\sigma_r = 4°$ $\sigma_n = 0$		135.6	88.9	276.9 (214.3)	182.8 (64.3)
$\sigma_r = 6°$ $\sigma_n = 0$		135.6	88.9	266.6 (204.2)	189.9 (68.9)

[a] numbers in parentheses are mean errors

TABLE 3.5.28

WIND SPEED INVERSION ERRORS USING

MICROWAVE MEASUREMENTS

• 662 meas over water clouds

		Regression		Kalman Filter	
		linear	nonlinear	EKF	IEKF
a priori		5.83 (0)	5.83 (0)	5.83 (0)	5.83 (0)
σ_r = 0.5°	σ_n = 0	5.17	2.02	5.89 (-1.35)	6.92 (-4.72)[a]
	σ_n = 1°		2.91	5.88 (-1.36)	7.18 (-4.80)
	σ_n = 5°		132.5	5.93 (-1.38)	8.02 (-3.69)
σ_r = 2°	σ_n = 0	5.17	2.02	5.85 (-1.24)	6.67 (-3.76)
	σ_n = 1°		2.91	5.85 -1.24	6.75 (-3.80)
	σ_n = 5°		132.5	5.89 (-1.27)	7.40 (-3.76)
σ_r = 4° σ_n = 0		5.17	2.02	5.78 (-1.01)	6.91 (-3.12)
σ_r = 6° σ_n = 0		5.17	2.02	5.72 (-0.79)	6.66 (-2.71)

[a] numbers in parentheses are mean errors

TABLE 3.5.29

CLOUD THICKNESS INVERSION ERRORS USING

MICROWAVE MEASUREMENTS

• 662 meas over water clouds

		Regression		Kalman Filter	
		linear	nonlinear	EKF	IEKF
a priori		12.20 (0)	12.20 (0)	12.20 (0)	12.20 (0)
σ_r = 0.5°	σ_n = 0	5.36	4.67	7.21 (-3.50)	8.46 (-1.13)[a]
	σ_n = 1°		4.75	7.27 (-3.49)	9.32 (-1.15)
	σ_n = 5°		7.30	8.89 (-3.45)	12.65 (-1.49)
σ_r = 2°	σ_n = 0	5.36	4.67	7.36 (-1.96)	7.70 (-0.43)
	σ_n = 1°		4.75	7.38 (-1.95)	7.61 (-0.47)
	σ_n = 5°		7.30	8.53 (-1.90)	9.52 (-0.37)
σ_r = 4° σ_n = 0		5.36	4.67	7.66 (0.28)	6.98 (-0.24)
σ_r = 6° σ_n = 0		5.36	4.67	7.82 (1.47)	7.05 (0.27)

[a] numbers in parentheses are mean errors

TABLE 3.5.30

CLOUD TOP HEIGHT INVERSION ERRORS

USING MICROWAVE MEASUREMENTS

• 662 meas over water clouds

		Regression		Kalman Filter	
		linear	nonlinear	EKF	IEKF
a priori		12.85 (0)	12.85 (0)	12.85 (0)	12.85 (0)
$\sigma_r = 0.5°$	$\sigma_n = 0$	10.76	7.31	15.10 (-4.40)	12.90 (-3.93)[a]
	$\sigma_n = 1°$		8.13	15.12 (-4.40)	16.15 (-3.76)
	$\sigma_n = 5°$		21.77	15.50 -4.37	25.4 (-3.0)
$\sigma_r = 2°$	$\sigma_n = 0$	10.76	7.31	15.04 (-4.84)	11.00 (-3.96)
	$\sigma_n = 1°$		8.13	15.06 (-4.83)	11.17 (-3.87)
	$\sigma_n = 5°$		21.77	15.40 (-4.81)	15.27 (-3.57)
$\sigma_r = 4°$ $\sigma_n = 0$		10.76	7.31	14.90 (-5.41)	10.80 (-3.30)
$\sigma_r = 6°$ $\sigma_n = 0$		10.76	7.31	14.71 (-5.54)	10.65 (-2.69)

[a] numbers in parentheses are mean errors

TABLE 3.5.31

WATER VAPOR INVERSION ERRORS

USING MICROWAVE MEASUREMENTS

• 662 meas over water clouds

		Regression		Kalman Filter	
		linear	nonlinear	EKF	IEKF
a priori		.539 (0)	.539 (0)	.539 (0)	.539 (0)
$\sigma_r = 0.5°$	$\sigma_n = 0$	.312	.204	.326 (-.188)	.327 (-.001)[a]
	$\sigma_n = 1°$		.277	.330 (-.189)	.386 (-.001)
	$\sigma_n = 5°$		8.13	.419 (-.188)	.678 (.023)
$\sigma_r = 2°$	$\sigma_n = 0$	.312	.204	.330 (-.10)	.325 (.050)
	$\sigma_n = 1°$		.277	.333 (-.097)	.327 (.049)
	$\sigma_n = 5°$		8.13	.392 (-.098)	.440 (.058)
$\sigma_r = 4°$ $\sigma_n = 0$		.312	.204	.352 (.038)	.332 (.047)
$\sigma_r = 6°$ $\sigma_n = 0$		.312	.204	.368 (.114)	.346 (.053)

[a] numbers in parentheses are mean errors

TABLE 3.5.32

RAIN RATE INVERSION ERRORS USING

MICROWAVE MEASUREMENTS

• 662 meas over water clouds

		Regression		Kalman Filter	
		linear	nonlinear	EKF	IEKF
a priori		6.80 (0)	6.80 (0)	6.80 (0)	6.80 (0)
$\sigma_r = 0.5°$	$\sigma_n = 0$	3.77	3.12	4.62 (-4.80)	4.80 (2.64)[a]
	$\sigma_n = 1°$		3.20	4.68 (-4.81)	5.03 (2.50)
	$\sigma_n = 5°$		236	6.01 (-4.86)	6.44 (0.94)
$\sigma_r = 2°$	$\sigma_n = 0$	3.77	3.12	4.27 (-3.44)	4.44 (2.30)
	$\sigma_n = 1°$		3.20	4.32 (-3.45)	4.48 (2.31)
	$\sigma_n = 5°$		236	5.31 (-3.50)	5.31 (2.09)
	$\sigma_r = 4°$ $\sigma_n = 0$	3.77	3.12	4.03 (-1.31)	4.28 (1.94)
	$\sigma_r = 6°$ $\sigma_n = 0$	3.77	3.12	4.13 (0.05)	4.21 (1.59)

[a] numbers in parentheses are mean errors

TABLE 3.5.33

REGRESSIONS FOR LWC WITH MISSING DATA

- RMS Errors
- noise-free

	All Meas μw	T_{10} out		T_{19} out		T_{37} out		T_{94} out	
Order		model	ave	model	ave	model	ave	model	ave
0	427.0								
1	138.0	138.0	138.0	162.1	427.0	138.0	138.0	138.0	138.0
2	114.0	114.0	114.0	149.3	427.0	114.0	114.0	114.0	114.0
3	107.5	107.5	107.5	107.8	427.0	107.5	107.5	107.5	107.5
4	100.2	102.4	284.6	104.1	187.3	101.9	271.7	100.2	100.2
5	95.6	100.6	262.8	100.2	174.2	97.8	250.6	95.6	95.6
6	95.1	99.7	227.2	95.5	171.1	96.3	241.2	95.1	95.4
7	89.1	94.6	208.1	89.6	214.1	90.1	129.9	89.1	92.1
8	85.8	90.5	171.7	86.9	341.9	86.1	88.0	85.8	87.9
9	84.5		207.8		352.2	84.8	86.0	84.5	87.9
10	82.7		329.7		643.0	83.3	85.0	82.7	87.5

TABLE 3.5.34

REGRESSIONS FOR WST WITH MISSING DATA

- RMS Errors
- noise-free

	All Meas μw	T_{10} out		T_{19} out		T_{37} out		T_{94} out	
Order		model	ave	model	ave	model	ave	model	ave
0	5.83								
1	5.45	5.61	5.83	5.45	5.45	5.45	5.45	5.45	5.45
2	3.07	4.10	21.54	3.58	20.03	3.07	3.07	3.07	3.07
3	2.89	3.90	20.09	3.33	14.64	2.94	5.36	2.89	2.89
4	2.80	2.83	20.86	3.21	15.93	2.83	4.89	2.80	2.80
5	2.74	2.77	21.58		17.55	2.78	4.13	2.74	2.74
6	2.62	2.73	25.83		23.09	2.66	3.87	2.62	2.62
7	2.42	2.53	26.22		21.43	2.52	7.06	2.55	2.89
8	2.38	2.44	28.55		25.29	2.41	5.71	2.44	3.00
9	2.32	2.42	29.57		27.03	2.40	5,69	2.36	3.06
10	2.11		28.66		27.61	2.38	14.40	2.12	2.67

TABLE 3.5.35

REGRESSIONS FOR CT WITH MISSING DATA

- RMS Errors
- noise-free

	All Meas μw	T_{10} out		T_{19} out		T_{37} out		T_{94} out	
Order		model	ave	model	ave	model	ave	model	ave
0	12.20								
1	5.52	5.52	5.52	5.52	5.52	5.90	13.75	5.58	6.13
2	5.34	5.34	5.34	5.34	5.34	5.36	13.78	5.41	5.80
3	5.27	5.27	5.27	5.29	5.34	5.21	13.83	5.32	5.74
4	4.88	5.18	5.97	5.12	5.64	5.04	15.29	5.11	5.51
5	4.82	5.11	8.25	5.02	9.34	4.95	11.80	5.07	5.35
6	4.74	4.90	10.18	4.95	7.31	4.90	12.28	5.05	5.37
7	4.67	4.80	9.96	4.75	11.94	4.79	8.70	4.93	5.35
8						4.78		4.78	
9									
10									

TABLE 3.5.36

REGRESSIONS FOR CH WITH MISSING DATA

- RMS Errors
- noise-free

	All Meas μw	T_{10} out		T_{19} out		T_{37} out		T_{94} out	
Order		model	ave	model	ave	model	ave	model	ave
0	12.85								
1	11.58	11.58	11.58	11.58	11.58	11.99	12.95	11.90	12.01
2	10.39	10.39	10.39	10.39	10.39	11.59	15.50	11.78	13.75
3	10.14	10.14	10.14	10.27	13.05	11.32	21.41	11.59	14.39
4	9.09	9.09	9.09	9.14	22.81	10.69	33.60		17.85
5	8.71	8.71	8.71	7.97	36.82	10.47	53.02		19.69
6	7.99	7.99	7.99	7.86	64.21	10.35	87.63		19.25
7	7.75	7.75	7.75	7.78	83.13	10.25	91.97		19.75
8	7.56		18.26	7.58	86.29	10.05	93.17		21.17
9	7.40		18.57		90.31		107.99		22.14
10	7.31		16.90		81.60		99.58		21.23

TABLE 3.5.37

REGRESSIONS FOR WV WITH MISSING DATA

- RMS Errors
- noise-free

	All Meas μw	T_{10} out		T_{19} out		T_{37} out		T_{94} out	
Order		model	ave	model	ave	model	ave	model	ave
0	.539								
1	.345	.345	.345	.367	.511	.345	.345	.349	.353
2	.327	.330	.548	.361	.839	.327	.327	.332	.338
3	.293	.329	.859	.333	1.670	.316	.615	.323	.338
4	.275	.315	1.377	.332	2.553	.314	.926	.265	.301
5	.226	.303	1.633		3.656	.309	1.505	.246	.290
6	.220	.290	1.650		3.613	.296	1.392	.245	.274
7	.220	.287	1.787		3.883	.291	1.509		.278
8	.216		1.816		3.937	.289	1.540		.283
9	.207		1.737		3.881		1.543		.285
10	.204		1.668		3.783		1.600		.285

TABLE 3.5.38

REGRESSIONS FOR R WITH MISSING DATA

- RMS Errors
- noise-free

	All Meas μw	T_{10} out		T_{19} out		T_{37} out		T_{94} out	
Order		model	ave	model	ave	model	ave	model	ave
0	6.80								
1	4.35	4.35	4.35	4.63	6.80	4.35	4.35	4.35	4.35
2	3.70	3.78	9.32	4.19	20.92	3.78	9.17	3.70	3.70
3	3.53	3.53	8.37	4.09	19.01	3.64	8.17	3.53	3.53
4	3.38	3.45	7.26	3.78	16.44	3.53	7.03	3.40	3.50
5	3.26	3.39	9.43	3.75	18.07	3.37	7.95	3.31	3.26
6	3.23	3.37	9.49	3.74	18.26	3.33	9.13	3.23	3.23
7	3.20	3.34	13.72	3.71	16.49	3.25	9.91	3.20	3.20
8	3.18		15.03	3.68	16.02		14.22	3.109	3.18
9	3.12		15.24		11.80		16.44	3.16	3.79
10									

TABLE 3.5.39 Mean Inversion Errors with Missing Data. Microwave Measurements over Water Clouds

· 662 meas.

· $\sigma_r = 2°$, $\sigma_n = 0$

		LWC	WST	CT	CH	WV	R
all μw channels	reg	0	0	0	0	0	0
	EKF	225.4	-1.24	-1.96	-4.84	-0.100	-3.44
	IEKF	76.2	-3.76	-0.43	-3.96	0.050	2.30
T_{10} out		-320.3	1.99	10.73	12.75	1.129	-8.17
		259.0	-2.70	5.06	-6.11	0.167	-2.76
		116.7	-4.92	1.47	-3.68	0.140	2.78
T_{19} out		-365.1	-0.01	-10.82	104.16	-.431	-9.81
		234.4	-0.77	4.92	-5.95	0.288	2.29
		102.0	-3.22	0.71	-3.26	0.103	2.19
T_{37} out		-2.53	14.22	-5.33	59.27	0.614	-14.14
		89.1	11.06	-2.87	-4.74	0.024	2.38
		36.3	-0.01	-0.13	-1.53	-0.026	0.12
T_{94} out		7.6	1.06	-0.38	0.54	0.119	0.38
		217.1	-0.79	-3.75	-13.21	0.017	-1.11
		58.7	0.12	0.34	0.41	-0.02	0.22

TABLE 3.5.40 RMS Inversion Errors with Missing Data. Microwave Measurements over Water Clouds

· 662 meas.

· $\sigma_r = 2°$, $\sigma_n = 0$

		LWC	WST	CT	CH	WV	R
apriori		427.0	5.83	12.20	12.85	0.539	6.80
all μw channels	reg	82.7	2.11	4.67	7.31	0.204	3.12
	EKF	287.3	5.85	7.36	15.04	0.330	4.27
	IEKF	235.1	6.67	7.70	11.00	0.325	4.44
T_{10} out		329.7	28.66	9.96	16.90	1.668	15.24
		294.3	6.01	9.08	14.77	0.371	4.33
		277.7	6.42	9.34	11.16	0.407	5.00
T_{19} out		643.0	27.61	11.94	81.60	3.783	11.80
		291.9	5.79	9.63	15.04	0.449	4.76
		244.1	6.53	8.67	10.67	0.403	5.18
T_{37} out		85.0	14.40	8.70	99.58	1.600	16.44
		209.3	5.62	7.17	15.03	0.365	4.76
		195.1	6.11	6.79	10.21	0.343	3.93
T_{94} out		87.5	2.67	5.35	21.23	0.285	3.26
		279.3	5.83	6.72	16.90	0.387	4.35
		243.9	5.84	7.75	13.74	0.372	4.82

TABLE 3.5.41

CRAMER-RAO BOUNDS FOR INVERSION ERRORS

• rms errors (662 meas)

	a priori	LWC	WST	DLT	CT	CH	WV	R
σ_r = 0.5°	10	20.3	0.30	2.09	0.57	0.67	0.23	0.66
	19	11.7	0.22	2.09	2.33	2.70	0.19	0.51
	37	3.48	0.28	2.09	0.70	0.40	0.24	1.23
	94	0.59	1.06	1.81	0.78	0.63	0.31	4.32
	10,19	8.79	0.18	2.09	0.55	0.60	0.12	0.34
	10,37	2.02	0.20	2.09	0.35	0.31	0.19	0.24
	10,94	0.59	0.28	1.19	0.37	0.36	0.21	0.26
	19,37	1.79	0.17	2.09	0.44	0.33	0.12	0.30
	19,94	0.59	0.22	1.33	0.45	0.45	0.13	0.31
	37,94	0.58	0.27	1.21	0.41	0.32	0.20	0.96
	10,19,37	1.66	0.15	2.09	0.32	0.29	0.09	0.19
	10,19,94	0.58	0.17	1.14	0.34	0.35	0.09	0.20
	10,37,94	0.56	0.19	1.06	0.28	0.26	0.15	0.23
	19,37,94	0.56	0.17	1.11	0.33	0.28	0.11	0.29
	10,19,37,94	0.55	0.14	1.04	0.26	0.24	0.09	0.19
σ_r = 1.0°	10	37.5	0.60	2.09	1.13	1.31	0.29	1.22
	19	20.3	0.44	2.09	3.83	4.35	0.27	0.98
	37	6.31	0.55	2.09	1.34	0.76	0.29	2.16
	94	1.18	2.01	2.00	1.35	0.96	0.32	4.32
	10,19	16.4	0.35	2.09	1.09	1.19	0.20	0.66
	10,37	3.71	0.40	2.09	0.69	0.60	0.26	0.47
	10,94	1.17	0.54	1.69	0.71	0.66	0.28	0.49
	19,37	3.56	0.35	2.09	0.88	0.66	0.20	0.56
	19,94	1.17	0.43	1.78	0.89	0.78	0.21	0.56
	37,94	1.14	0.52	1.71	0.79	0.58	0.27	1.61
	10,19,37	3.29	0.30	2.09	0.64	0.57	0.17	0.38
	10,19,94	1.16	0.34	1.66	0.66	0.64	0.17	0.38
	10,37,94	1.11	0.38	1.60	0.55	0.48	0.23	0.45
	19,37,94	1.11	0.33	1.63	0.64	0.52	0.19	0.53
	10,19,37,94	1.10	0.29	1.58	0.51	0.46	0.16	0.36

NOTE: The Cramer-Rao bound is the absolute lower bound on the inversion rms error for <u>any</u> (linear or nonlinear) processor.

TABLE 3.5.42

CRAMER-RAO BOUNDS FOR INVERSION ERRORS

• rms errors (662 meas)

	a priori	LWC	WST	DLT	CT	CH	WV	R
σ_r = 2°	10	61.5	1.17	2.09	2.16	2.45	0.32	1.98
	19	33.6	0.87	2.09	5.27	6.03	0.31	1.77
	37	10.6	1.07	2.09	2.43	1.47	0.32	3.21
	94	2.3	3.42	2.06	2.38	1.63	0.33	4.32
	10,19	28.5	0.70	2.09	2.08	2.29	0.28	1.24
	10,37	6.9	0.78	2.09	1.36	1.18	0.31	0.90
	10,94	2.3	1.04	1.97	1.39	1.22	0.31	0.93
	19,37	7.0	0.68	2.09	1.70	1.32	0.28	0.98
	19,94	2.3	0.84	1.99	1.72	1.39	0.28	0.95
	37,94	2.2	0.99	1.97	1.50	1.06	0.31	2.56
	10,19,37	6.4	0.58	2.09	1.26	1.14	0.25	0.71
	10,19,94	2.3	0.67	1.95	1.29	1.19	0.25	0.70
	10,37,94	2.2	0.73	1.93	1.09	0.92	0.29	0.87
	19,37,94	2.2	0.65	1.94	1.25	0.98	0.27	0.91
	10,19,37,94	2.2	0.56	1.92	1.02	0.89	0.24	0.67
σ_r = 5°	10	91.9	2.62	2.09	4.41	5.12	0.33	2.82
	19	57.0	2.03	2.09	6.32	8.13	0.34	2.96
	37	19.6	2.45	2.09	4.42	3.51	0.34	3.99
	94	5.8	5.06	2.08	4.32	3.53	0.34	4.32
	10,19	52.9	1.67	2.09	4.20	4.84	0.32	2.38
	10,37	16.3	1.85	2.09	3.07	2.84	0.33	2.01
	10,94	5.8	2.35	2.07	3.09	2.84	0.33	2.04
	19,37	16.6	1.62	2.09	3.62	3.18	0.32	1.99
	19,94	5.8	1.92	2.07	3.64	3.18	0.33	1.93
	37,94	5.5	2.14	2.07	3.20	2.50	0.33	3.73
	10,19,37	15.4	1.41	2.09	2.89	2.75	0.31	1.57
	10,19,94	5.7	1.58	2.07	2.93	2.78	0.31	1.51
	10,37,94	5.4	1.71	2.06	2.52	2.20	0.33	1.93
	19,37,94	5.5	1.52	2.07	2.82	2.36	0.32	1.84
	10,19,37,94	5.4	1.34	2.06	2.38	2.14	0.31	1.46

IV. INVERSIONS FOR CONTINUOUS-VALUED PARAMETERS USING VISIBLE, NEAR IR AND IR MEASUREMENTS

Cloud identification using visible, near infrared and infrared data is performed using the same techniques as used for microwave data. The first step consists of forward simulations to compute intensities and/or brightness temperatures as a function of cloud and atmospheric parameters. The simulations are performed in two steps, as shown in Figure 4.1. The program RASP is used to compute the extinction coefficient and single scattering albedo for a given cloud configuration. A modified version of the LOWTRAN 3B program is then used to solve the radiative transfer problem.

The basic assumptions and conditions used in the simulations are now given. A midlatitude Spring-Fall standard atmosphere was used. Cloud-free visibility was 23 km. The simulations were run over ocean for two different data sets, with the surface reflections used given in Table 4.1. The sun was assumed directly overhead. Additional inputs required are the values of solar flux and the index of refraction; the values of solar flux are given in Table 4.1 and the index of refraction of water at the selected wavelengths is given in Table 4.2. In the table, r_c, C_1, C_2 are the shape parameters, respectively of the analytical distribution of Deirmendjian (1964) for characterizing the cloud layers.

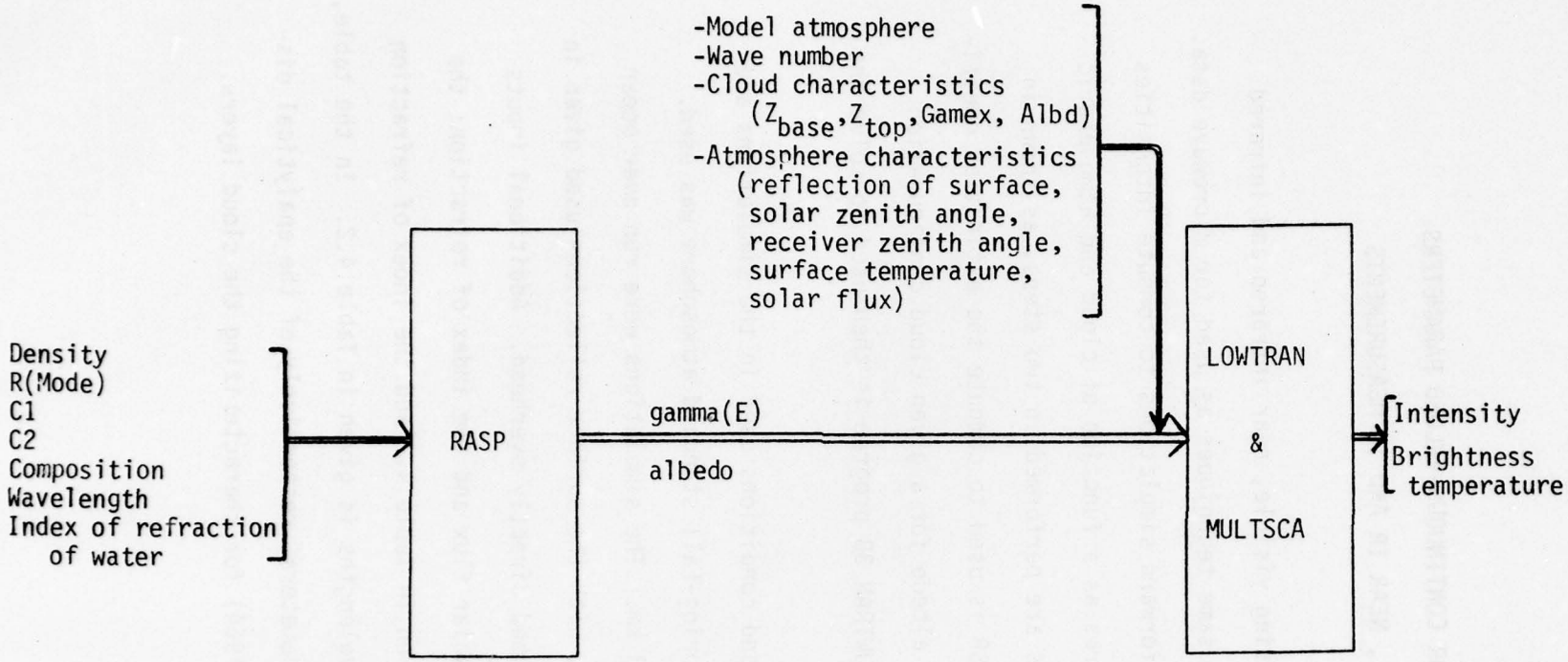


Figure 4.1 Forward Simulation Procedure

TABLE 4.1

INPUT DATA USED FOR LOWTRAN & MULTSCA

WAVELENGTH (microns)	WAVE NUMBER (cm^{-1})	SOLAR FLUX ($erg\text{-}sec^{-1}\text{-}cm^{-2}\text{-}strd^{-1}\text{-}cm^{-1}$)	REFLECTION OF SURFACE
.55	18180	6.201 E+9	.05
.725	13790	4.520 E+9	.025
1.0	10000	2.324 E+9	.02
1.6	6250	6.366 E+8	.02
2.1	4760	2.706 E+8	.02
3.8	2630	3.406 E+7	.02
6.7	1490	0.	0.
10.5	950	0.	0.
11.5	870	0.	0.
12.5	800	0.	0.

Z_{base}-Z_{top} (m)
2400-2900
2400-4900
4400-4900
330- 660
660-1320

Solar zenith angle: overhead sun

Receiver zenith angle: overhead

Surface temperature: 288.6°K

TABLE 4.2

INPUT DATA USED FOR RASP

Density = .1, .25, .5, .75, 1., 1.25, 1.5, 1.75, 2. (g/m^3)

r_c = 10 microns, C_1 = 6, C_2 = 0.5

Composition = liquid

WAVELENGTH (cm)	INDEX OF REFRACTION OF WATER	
	R_e	I_m
.55 E-4	1.333 E+0,	-1.960 E-9
.725 E-4	1.331 E+0,	-3.280 E-8
1.0 E-4	1.326 E+0,	-3.040 E-6
1.6 E-4	1.316 E+0,	-1.0 E-4
2.1 E-4	1.3 E+0,	-8.0 E-4
3.8 E-4	1.366 E+0,	-3.70 E-4
6.7 E-4	1.330 E+0,	-3.70 E-2
10.5 E-4	1.179 E+0,	-6.74 E-2
11.5 E-4	1.126 E+0,	- .142 E+0
12.5 E-4	1.123 E+0,	- .259 E+0

4.1 Inversions For Visible/IR Dataset

A series of forward simulations was made to generate models for predicting intensity and brightness temperature. The atmospheric conditions used in these simulations were derived from microwave dataset #3 by eliminating cases containing variations in wind speed and rainfall rate, since these variables are expected to have little effect on the observed intensities. A summary of the distribution of parameters is given in Table 4.1.1. The values of the simulated intensity for the visible and near IR channels (.55μ - 3.8μ) were then scaled by the incident solar flux to give an equivalent albedo. This section and those that follow will use the notation $A\cdot\lambda$ for albedo at wavelength λ and $I\cdot\lambda$ for intensity at wavelength λ.

A series of forward models for predicting the observation given the meteorological parameters were developed. The independent variables for these regressions were nonlinear functions of the parameters LWC, CH and CT. Also used in these models were powers of the cloud optical depth (OD). The later parameter is basically proportional to the LWC. The dependent variable was either albedo or observed radiance.

The results of these forward models are given in Tables 4.1.2 through 4.1.11. In the regression models for intensity ($\lambda = 6.7\mu - 12.5\mu$), the units of the dependent variables are 10^7 erg-sec^{-1}-cm^{-2}-str^{-1}-cm^{-1}.

It may be seen that the models in the visible range are quite nonlinear in LWC (or its equivalent OD) with other parameters entering only later in the regression. For the near-IR channels, the forward

models begin to place more emphasis on CH as a predictor. In the models for the thermal IR channels, the intensity is a strong function of CH with little dependence on LWC.

Inversion Using Linear Regression

A series of noise-free inversions using linear regression models was performed. The results are shown in Tables 4.1.12 - 4.1.14, in which the models and statistical performance are shown for estimating LWC, CT and CH. Cloud top height estimation performance was the best with an expected strong dependence on a thermal IR channel. Liquid water content and cloud thickness were estimated less well with the 3.8 μ channel being the best linear predictor. It should be noted that the strong nonlinear dependence of visible channel albedo on LWC is not exploited by the linear regression scheme of this experiment.

TABLE 4.1.1

VISIBLE AND INFRARED

DATA SET

190 CASES

LWC ($gm/cm^2 \times 10^3$)

	6(6)60	26(26)260	4.3(4.3)43	7.6(7.6)76	286(26)520
#	110	30	10	10	30

DLT (°K)

	-5	0	5
#	20	150	20

CT (m)

	330	500	660	2500
#	10	110	10	60

CH (m)

	660	1320	1500	2900	4900
#	10	10	10	50	110

WV (gm/cm^2)

	.89	1.02	1.38	1.62	1.70	1.82	1.93	2.29	2.34	2.44
#	10	10	20	30	30	20	30	10	10	20

TABLE 4.1.2

REGRESSION MODEL FOR $A_{.55}$

Variable	Coefficient	F Value
$LWC^2 \times CH^2$	4.195×10^{-10}	7.14
OD	-7.120×10^{-3}	1199.41
OD^2	9.866×10^{-5}	627.38
OD^3	-6.637×10^{-7}	607.18
OD^4	2.004×10^{-9}	488.03
OD^5	-2.260×10^{-12}	408.66
1/OD	-1.882	5413.62
$1/OD^4$	2.837×10^{1}	710.00
$1/OD^5$	$-4.570\ \ 10^{1}$	585.85

R^2 .999

Intercept Estimate 1.096

Error Standard Deviation 2.717×10^{-3}

TABLE 4.1.3

REGRESSION MODEL FOR $A_{.725}$

Variable	Coefficient	F Value
$LWC^2 \times CH$	-2.380×10^{-6}	65.86
$1/(LWC^2 \times CH)$	-1.967×10^{1}	86.88
$LWC^2 \times CH^2$	5.943×10^{-8}	77.39
$LWC^3 \times CH^2$	-6.224×10^{-8}	84.80
OD	-6.806×10^{-3}	175.65
OD^3	6.230×10^{-6}	83.40
OD^5	7.064×10^{-13}	105.41
1/OD	-1.753	1579.66
$1/OD^4$	9.414	199.61

R^2 .987

Intercept Estimate 1.093

Error Standard Deviation 9.724×10^{-3}

TABLE 4.1.4

REGRESSION MODEL FOR A_1.

Variable	Coefficient	F Value
$LWC^2 \times CH^2$	5.838×10^{-9}	580.14
$LWC^3 \times CH^2$	-1.754×10^{-10}	253.88
OD	-3.778×10^{-3}	881.87
OD^3	1.612×10^{-6}	242.53
OD^5	3.297×10^{-13}	361.90
1/OD	-1.349	1983.33
$1/OD^5$	5.203	730.15

R^2 .998

Intercept Estimate 1.013

Error Standard Deviation 3.540×10^{-3}

TABLE 4.1.5

REGRESSION MODEL FOR $A_{1.6}$

Variable	Coefficient	F Value
WV	-2.721×10^{-4}	5.538
$1/(LWC^2 \times CH)$	5.369	969.36
$LWC^2 \times CH^2$	-2.972×10^{-9}	1083.49
$LWC^3 \times CH^2$	5.653×10^{-11}	1174.10
OD	7.900×10^{-3}	1533.67
OD^2	-9.272×10^{-5}	1290.79
OD^4	-2.777×10^{-9}	1086.12
OD^5	3.499×10^{-12}	1021.68
1/OD	3.577	2760.95
$1/OD^2$	-2.275×10^{1}	3650.22
$1/OD^3$	5.082×10^{1}	3218.55
$1/OD^4$	-3.982×10^{1}	2844.19

R^2 .999

Intercept Estimate 4.980×10^{-1}

Error Standard Deviation 6.411×10^{-4}

TABLE 4.1.6

REGRESSION MODEL FOR $A_{2.1}$

Variable	Coefficient	F Value
CH	-2.044×10^{-5}	2287.19
LWC × CT	-2.698×10^{-6}	9422.25
LWC × CH	4.113×10^{-7}	918.70
$LWC^2 \times CH$	5.425×10^{-8}	7621.10
OD^2	-2.372×10^{-6}	7646.43
OD^3	-3.523×10^{-6}	2319.69
1/OD	-2.046×10^{-1}	16163.84
$1/OD^2$	2.061	25235.01
$1/OD^3$	-5.995	24528.71
$1/OD^4$	-1.255×10^{1}	19228.44
$1/OD^5$	3.609×10^{1}	67999.58

R^2 .999

Intercept Estimate 4.858×10^{-1}

Error Standard Deviation 3.929×10^{-5}

TABLE 4.1.7

REGRESSION MODEL FOR $A_{3.8}$

Variable	Coefficient	F Value
CH	-6.234×10^{-4}	403.15
LWC × CH	-1.685×10^{-5}	635.04
$LWC^2 \times CH$	2.786×10^{-7}	403.49
OD^2	-2.095×10^{-5}	225.33
OD^3	-1.183×10^{-8}	51.04
OD^5	1.716×10^{-14}	13.03

R^2 .990

Intercept Estimate 3.284×10^{-1}

Error Standard Deviation 3.377×10^{-3}

TABLE 4.1.8

REGRESSION MODEL FOR $I_{6.7}$

Variable	Coefficient	F Value
CH	-4.007×10^{-2}	1960.46
1/OD	5.995×10^{-1}	11.23

R^2 .923

Intercept Estimate 3.738

Error Standard Deviation 1.705×10^{-1}

TABLE 4.1.9

REGRESSION MODEL FOR $I_{10.5}$

Variable	Coefficient	F Value
CH	-5.830×10^{-2}	2731.59
1/OD	1.338	34.78

R^2 .945

Intercept Estimate 6.641

Error Standard Deviation 2.101×10^{-1}

TABLE 4.1.10

REGRESSION MODEL FOR $I_{11.5}$

Variable	Coefficient	F Value
CH	-5.520×10^{-2}	2694.38
1./OD	1.083	32.95

R^2 .944

Intercept Estimate 6.729

Error Standard Deviation 2.003×10^{-1}

TABLE 4.1.11

REGRESSION MODEL FOR $I_{12.5}$

Variable	Coefficient	F Value
CH	-4.957×10^{-2}	2719.44
1./OD	1.032	34.90

R^2 .945

Intercept Estimate 6.357

Error Standard Deviation 1.790×10^{-1}

AD-A077 873 SCIENTIFIC SYSTEMS INC CAMBRIDGE MA F/G 4/2
MULTISPECTRAL CLOUD IDENTIFICATION STUDY.(U)
SEP 78 D E GUSTAFSON , W H LEDSHAM F19628-77-C-0208
UNCLASSIFIED AFGL-TR-78-0280 NL

3 OF 3
ADA
077873

END
DATE
FILMED
1 -80
DDC

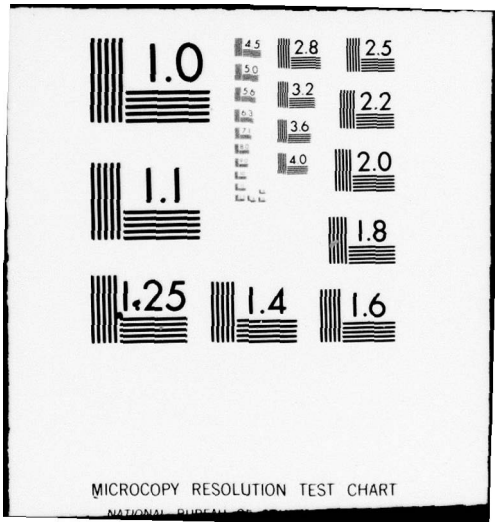

1.0
1.1
1.25
1.4
1.6
1.8
2.0
2.2
2.5
2.8
3.2
3.6
4.0
4.5
5.0
5.6
6.3
7.1
8.0
9.0
MICROCOPY RESOLUTION TEST CHART

TABLE 4.1.12

REGRESSION MODEL FOR LWC

Variable	Coefficient	F Value
$A_{1.6}$	-1.175×10^{3}	19.23
$A_{2.1}$	4.275×10^{3}	11.49
$A_{3.8}$	-5.802×10^{3}	926.69
$I_{10.5}$	3.108×10^{-4}	82.51
$I_{11.5}$	-3.189×10^{-4}	78.79

R^2 .875

Intercept Estimate 1.502

Error Standard Deviation 4.900×10^{1}

TABLE 4.1.13

REGRESSION MODEL FOR CT

(CT in m × 100)

Variable	Coefficient	F Value
$A_{.725}$	-4.820×10^{1}	11.80
$A_{1.6}$	6.650×10^{1}	8.01
$A_{3.8}$	-3.367×10^{2}	504.18
$I_{6.7}$	-7.926×10^{-7}	4.00
$I_{12.5}$	1.034×10^{-6}	9.60

R^2 .819

Intercept Estimate 6.637×10^{1}

Error Standard Deviation 4.029

TABLE 4.1.14

REGRESSION MODEL FOR CH

(CH in m × 100)

Variable	Coefficient	F Value
$A_{1.0}$	-1.581×10^{1}	22.73
$I_{10.5}$	-1.592×10^{-6}	2913.11

R^2 .942

Intercept Estimate 1.231×10^{2}

Error Standard Deviation 3.471

V. CLASSIFICATION OF DISCRETE VALUED PARAMETERS USING VISIBLE AND I.R. DATA. DISCRIMINATION OF ICE CLOUDS, WATER CLOUDS, CLEAR COLUMN WITH SNOW, CLEAR COLUMN WITH LOAM

5.1 Introduction

The nature of the problem addressed by this section is the one of automatic scene classification given a set of measurements in a spectral interval. Specifically, the problem at hand is to classify the scene into several categories: (1) a glaciated cloud or ice cloud, (2) a water cloud, (3) a clear column with snow or (4) a clear column with bare ground.

The spectral interval of interest for the purposes of this classification is the specified wavelengths in the band from .55 μ to 12.6 μ. The wavelengths in the microwave are excluded from consideration. Examination of brightness temperatures as a function of particle size, snow depth, etc. compiled for a model proposed by Chang et al. (1976) show that a meter of new snow is virtually indistinguishable from a bare surface, while 10 cm of snow subjected to repeated freeze thaw cycles responds quite noticeably. In general, Chang et al., state that the penetration depth of most snows range from 10 to 100 wavelengths. Due to this variability, the microwave channels will be excluded.

The approach that will be taken in the scene classification is the non-parametric approach of Friedman (1977). This approach has several desirable properties:

(i) The procedure is computationally simple for both training and classification phases. This is of great importance in the present problem, due to its high dimensionality.

(ii) The decision rule that results from the training phase is asymptotically Bayes' risk efficient as the number of training samples increases without limit.

(iii) Additional features may be easily added prior to the training phase. These may be linear or nonlinear. Noninformative features are simply ignored in the training phase. However, any of the additional features which are informative are used, and become part of the classification algorithm.

The discrimination experiment performed using the Friedman Tree approach entailed the discrimination of four scene categories: clear columns over snow, clear columns over loam, ice clouds over either background and water clouds over either background. The results of the experiment indicate that there exist robust features to discriminate clear columns and water clouds. Ice clouds are also usually discriminated, but the partitioning trees are more complex.

5.2 The Discrimination Experiment Data Base

The experimental data base used in both discrimination experiments was constructed using simulations from the modified LOWTRAN 3B code provided by Environmental Research and Technology Inc. (ERT). As presently implemented, this code provides the ability to compute the observed radiance or brightness temperature from a user specified atmosphere over a surface with a user specified reflectivity. A single cloud layer with a user specified absorption and single scattering albedo may be inserted in atmosphere. These two cloud quantities were computed for a cloud phase and drop size distribution using the ERT program RASP.

The data base was constructed from cloud thickness, phase and altitudes in a manner expected to span the single layer cloud types that might be encountered in mid-latitude and sub-arctic winter. Due to current limitations in the simulation software, multi-layer or multi-phase cloud types could not be simulated.

All cloud types were simulated over both a snow surface and a loam surface. The liquid phase clouds were simulated using the 45° winter supplemental atmcsphere internal to LOWTRAN. The ice phase clouds were simulated using both the LOWTRAN 60° winter and 45° winter supplemental atmospheres.

In the simulations that used the 45° latitude atmosphere, the entire simulation was run using a surface temperature of 265° K and again using a surface temperature of 272.15° K. For the 60° winter atmosphere, the surface temperatures used were 260° K and 255° K. The exception to this surface temperature practice were the simulated clear columns. For these cases, the surface temperature was varied from 255° K to 260° K in 0.5° K increments for the 60° winter atmosphere and 265° K to 271.5° K and 272.15°K in 0.5° K increments for the 45° winter atmosphere.

A summary of the reflectivities used for the loam and snow surfaces is given in Table 5-1. The figures for snow were abstracted from Valley (1966) and O'Brien and Munis (1974).

The reflectivity for 3.8 μm was extrapolated. The reflectivity values for loam were taken or extrapolated from Valley (1965). The values for the reflectivity at wavelengths greater than 4 μm was taken to be zero for both surface types.

TABLE 5-1

ASSUMED SURFACE REFLECTIVITES

WAVELENGTH (MICRONS)	SNOW REFLECTIVITY	LOAM REFLECTIVITY
.5501	0.737	0.0875
.7252	0.775	0.1875
1.0000	0.775	0.3000
1.6000	0.070	0.3000
2.1008	0.050	0.3000
3.8023	0.050	0.3000
6.7114	0.0	0.0
10.5263	0.0	0.0
11.4943	0.0	0.0
12.5000	0.0	0.0

The values of the solar flux were taken to be those given earlier in this report. It should be noted that these values will not be exactly correct for these simulations due to the difference in the inclination of the earth's axis between spring and winter and 45° and 60° latitude. For those cases in which the discriminating features are ratios of observed radiances in either the visible or near IR, this use of 45° spring-fall flux will be transparent. For other features it will be a simple shift in the cut value due to the monotone properties of the Friedman Tree algorithm.

A summary of the cloud altitudes for the liquid phase clouds simulated is given in Table 5-2. The liquid water densities for these clouds were simulated at .25 g/m^3, .5 g/m^3, .75 g/m^3 and 1 g/m^3 for all clouds and 1.25 g/m^3, 1.5 g/m^3, 1.75 g/m^3, and 2.0 g/m^3 for those clouds thicker than 1 km.

A summary of the cloud altitudes for the ice phase clouds is given in Table 5-3. These clouds were simulated at .025 g/m^3, .05 g/m^3, .1 g/m^3 and .2 g/m^3.

The computed absorption and single scattering albedo for a liquid cloud with a density of 1 g/m^3 and an ice cloud with a density of .1 g/m^3 are given in Tables 5-4 and 5-5.

The total number of simulations run was 599. Of these 74 were clear columns, 188 were water clouds and 337 were ice clouds.

TABLE 5-2

WATER CLOUD ALTITUDES

USED IN THE SIMULATION

BASE (METERS)	TOP (METERS)
500	1000
1000	1500
1500	2000
2000	2500
2500	3000
3000	3500
3500	4000
1000	2000
2000	3000
2500	3500
3000	4000
1000	3000
1500	3500
2000	4000

TABLE 5-3

ICE CLOUD ALTITUDES

USED IN THE SIMULATION

BASE (METERS)	TOP (METERS)
6000	6200
4000	4500
6000	6500
7000	7500
8000	8500
2000	3000
3000	4000
4000	5000
5000	6000
6000	7000
7000	8000
8000	9000
3000	5000

TABLE 5-4

WATER CLOUD CHARACTERISTICS

USED IN THE SIMULATION

FOR A DENSITY OF 1 g/m^3

WAVELENGTH (MICRONS)	EXTINCTION NEPERS/CM	SINGLE SCATTERING ALBEDO
.55	6.160×10^{-4}	1.00000
.72	6.164×10^{-4}	1.00000
1.00	6.252×10^{-4}	.99972
1.60	6.468×10^{-4}	.99159
2.10	7.172×10^{-4}	.94251
3.8	7.108×10^{-4}	.82479
6.7	7.200×10^{-4}	.55901
10.0	7.408×10^{-4}	.56201
11.5	6.460×10^{-4}	.46718
12.5	6.592×10^{-4}	.46395

TABLE 5-5

ICE CLOUD CHARACTERISTICS

USED IN THE SIMULATIONS

FOR A DENSITY OF .1 g/m^3

WAVELENGTH (MICRONS)	EXTINCTION NEPERS/CM	SINGLE SCATTERING ALBEDO
.55	1.540×10^{-5}	1.00000
.72	1.541×10^{-5}	1.00000
1.00	1.563×10^{-5}	0.99972
1.60	1.612×10^{-5}	0.99139
2.10	1.807×10^{-5}	0.93875
3.8	1.737×10^{-5}	0.77424
6.7	1.668×10^{-5}	0.50000
10.0	1.668×10^{-5}	0.52100
11.5	1.682×10^{-5}	0.50332
12.5	1.651×10^{-5}	0.50228

5.3 Partitioning Results

Using the simulations of the database, a single major experiment was performed. The classes to be discriminated were clear columns over loam, clear columns over snow, liquid clouds and ice clouds. The purpose of this experiment was twofold. First, it was performed in support of the inversion experiments in earlier sections of this report. These experiments rely on a clear column, water cloud, ice cloud discrimination for their choice of algorithm. The second purpose of this experiment was strictly meteorological as the presence or absence of clouds and their phase are important parameters in their own right. In the experiment, the sensitivity of the partitioning to the simulation accuracy was explored by repeating the experiments with 0.1, 0.5, 1.0, 5.0 and 10.0 percent multiplicative pseudorandom noise from a normal distribution on the data.

The features used in the discrimination experiments were the observed radiances at the visible and near IR wavelengths, brightness temperatures at the thermal wavelengths and the ratios between these variables. As before the radiance used in the experiment was erg-sec^{-1}-cm^{-2}-strd^{-1}-cm^{-1}. The monotone properties of the Friedman Tree algorithm guarantee that this choice of feature space will be fairly extensive. For the noiseless case, the use of any other monotone function of the observed radiances such as a Albedo would yield equivalent partitioning rules.

The partitioning trees for the experiment using 1.0 percent noise are given in Figs. 5.1 - 5.4. The trees for 1 percent noise rather than other values are given for the results of this experiment for reasons of robustness.

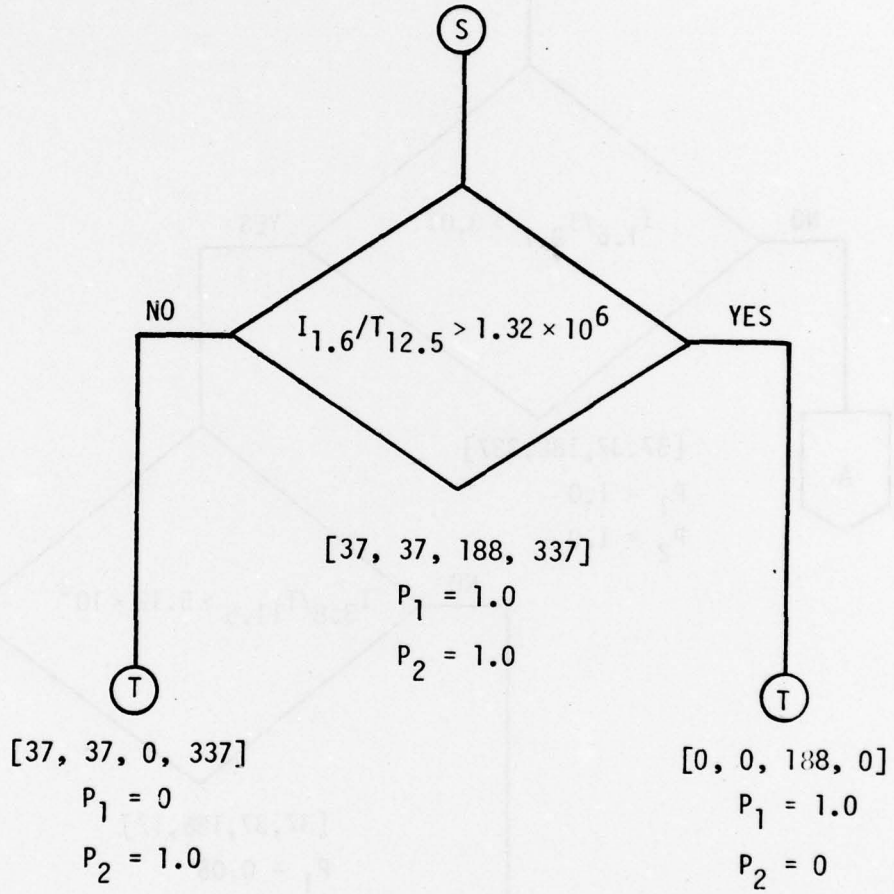


Figure 5.1. Decision Tree for Water Clouds
1.0% Multiplicative Noise

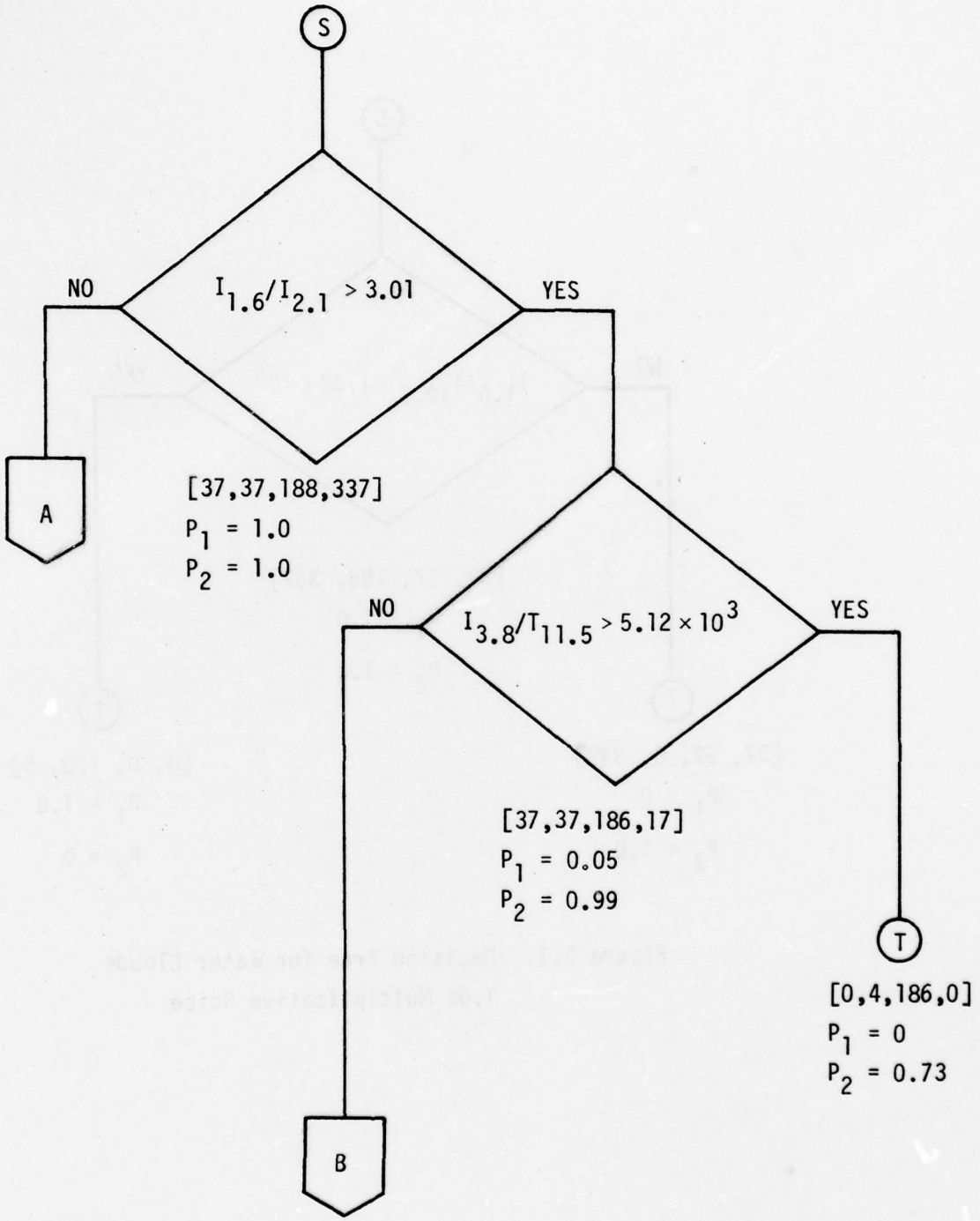


Figure 5.2. Decision Tree for Ice Clouds
1.0% Multiplicative Noise

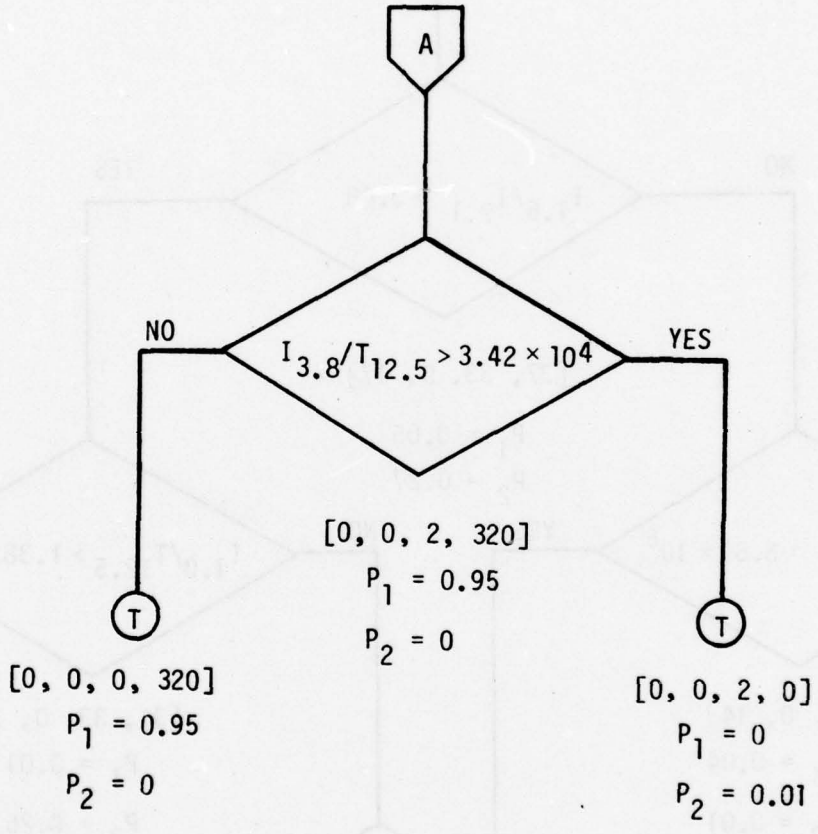


Figure 5.2 (Continued)

Decision Tree for Ice Clouds 1.0% Multiplicative Noise

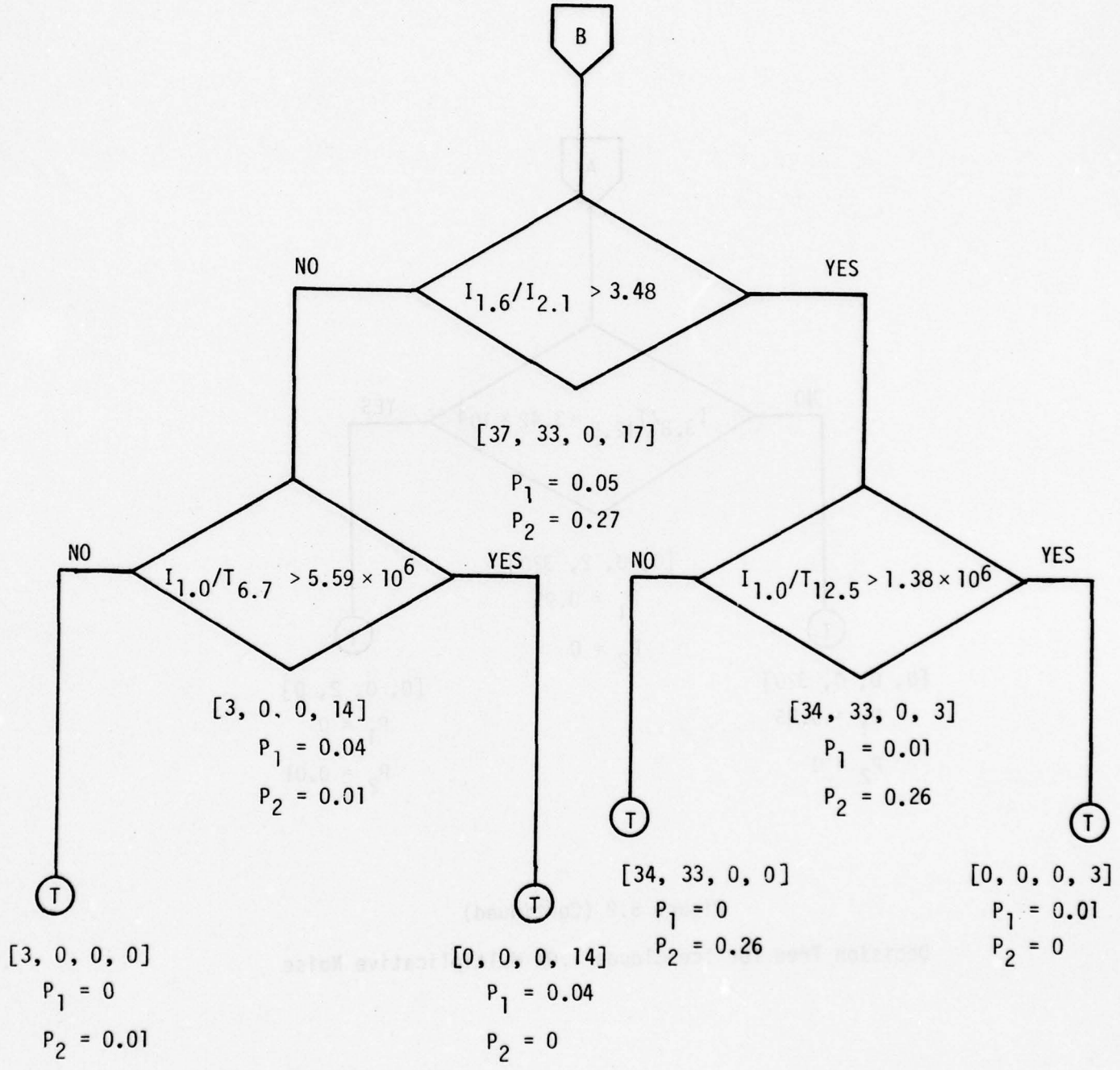


Figure 5.2 (Continued)

Decision Tree for Ice Clouds 1.0% Multiplicative Noise

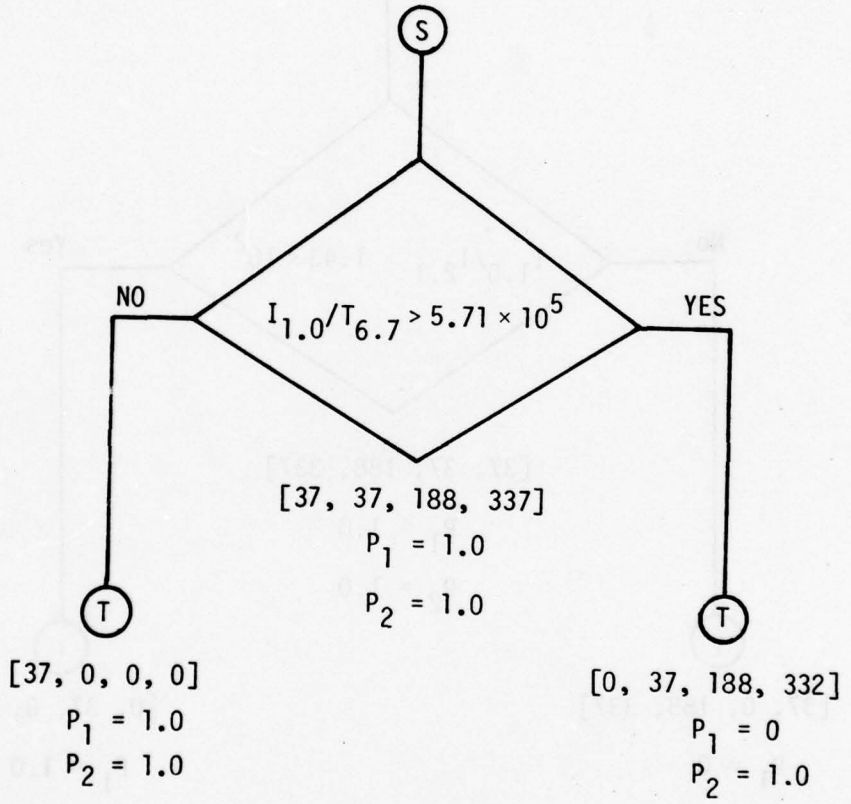


Figure 5.3. Decision Tree for Clear Columns Over Loam
1.0 % Multiplicative Noise

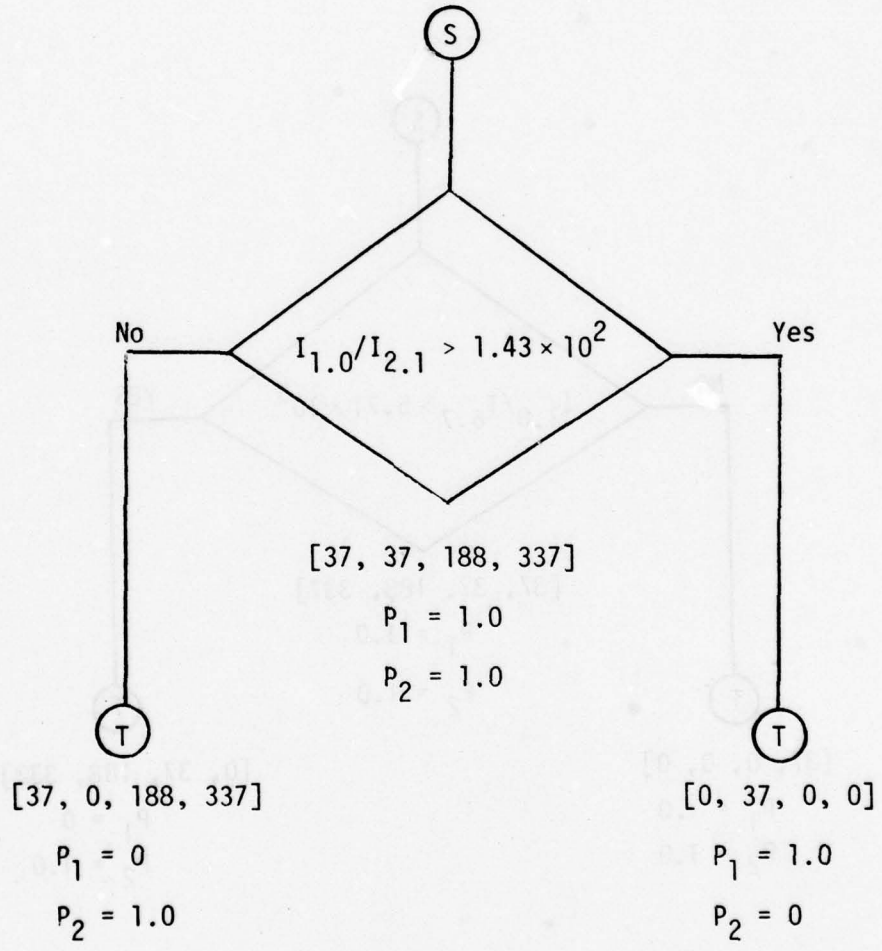


Fig. 5.4. Decision Tree For Clear Columns Over Snow
1.0% Multiplicative Noise

Decision Tree for Water Clouds

No Multiplicative Noise

Table 5.6

NODE	DISCRIMINATING FEATURE	CUT VALUE	MAXIMUM KS DISTANCE	NEXT NODE IF LESS THAN OR EQUAL	NEXT NODE IF GREATER THAN	TOTAL POPULATION AT NODE	MEMBERS IN CLASS	MEMBERS NOT IN CLASS	P_1	P_2	CLEAR COLUMNS OVER LOAM	CLEAR COLUMNS OVER SNOW	WATER CLOUDS	ICE CLOUDS
1	$T_{10.5}/T_{12.5}$	9.96×10^{-1}	1.00	3	2	599	188	411	1.0	1.0	37	37	188	337
2	Terminal	--	--	-	-	411	0	411	0.0	1.0	37	37	0	337
3	Terminal	--	--	-	-	188	188	0	1.0	0.0	0	0	188	0

Decision Tree for Ice Clouds

No Multiplicative Noise

Table 5.7

NODE	DISCRIMINATING FEATURE	CUT VALUE	MAXIMUM KS DISTANCE	NEXT NODE IF LESS THAN OR EQUAL	NEXT NODE IF GREATER THAN	TOTAL POPULATION AT NODE	MEMBERS IN CLASS	MEMBERS NOT IN CLASS	P_1	P_2	CLEAR COLUMNS OVER LOAM	CLEAR COLUMNS OVER SNOW	WATER CLOUDS	ICE CLOUDS
1	$I_{1.6}/I_{2.1}$	3.02	0.953	3	2	599	337	262	1.0	1.0	37	37	188	337
2	$T_{10.5}/T_{12.5}$	9.97×10^{-1}	0.870	5	4	278	16	262	0.05	1.0	37	37	188	16
3	Terminal	--	--	-	-	321	321	0	0.95	0.0	0	0	0	321
4	$T_{10.5}$	2.66×10^{2}	1.0	7	6	50	16	34	0.05	0.13	17	17	0	16
5	Terminal	--	--	-	-	228	0	228	0.0	0.87	20	20	188	0
6	Terminal	--	--	-	-	34	0	34	0.13		17	17	0	0
7	Terminal	--	--	-	-	16	16	0	0.05		0	0	0	16

Decision Tree for Clear Columns Over Loam

No Multiplicative Noise

Table 5.8

NODE	DISCRIMINATING FEATURE	CUT VALUE	MAXIMUM KS DISTANCE	NEXT NODE IF LESS THAN OR EQUAL	NEXT NODE IF GREATER THAN	TOTAL POPULATION AT NODE	MEMBERS IN CLASS	MEMBERS NOT IN CLASS	P_1	P_2	CLEAR COLUMNS OVER LOAM	CLEAR COLUMNS OVER SNOW	WATER CLOUDS	ICE CLOUDS
1	$I_{1.0}/T_{12.5}$	5.70×10^5	1.00	3	2	599	37	562	1.0	1.0	37	37	188	337
2	Terminal	--	--	-	-	562	0	562	0.0	1.0	0	37	188	337
3	Terminal	--	--	-	-	37	37	0	1.0	0.0	37	0	0	0

Decision Tree for Clear Columns Over Snow

No Multiplicative Noise

Table 5.9

NODE	DISCRIMINATING FEATURE	CUT VALUE	MAXIMUM KS DISTANCE	NEXT NODE IF LESS THAN OR EQUAL	NEXT NODE IF GREATER THAN	TOTAL POPULATION AT NODE	MEMBERS IN CLASS	MEMBERS NOT IN CLASS	P_1	P_2	CLEAR COLUMNS OVER LOAM	CLEAR COLUMNS OVER SNOW	WATER CLOUDS	ICE CLOUDS
1	$I_{1.0}/I_{2.1}$	1.43×10^2	1.00	3	2	599	37	562	1.0	1.0	37	37	188	337
2	Terminal	--	--	-	-	37	37	0	1.0	0.0	0	37	0	0
3	Terminal	--	--	-	-	562	0	562	0.0	1.0	37	0	188	337

Decision Tree for Water Clouds

0.1% Multiplicative Noise

Table 5.10

NODE	DISCRIMINATING FEATURE	CUT VALUE	MAXIMUM KS DISTANCE	NEXT NODE IF LESS THAN OR EQUAL	NEXT NODE IF GREATER THAN	TOTAL POPULATION AT NODE	MEMBERS IN CLASS	MEMBERS NOT IN CLASS	P_1	P_2	CLEAR COLUMNS OVER LOAM	CLEAR COLUMNS OVER SNOW	WATER CLOUDS	ICE CLOUDS
1	$I_{1.6}/T_{12.5}$	1.31×10^6	1.00	2	3	599	188	411	1.0	1.0	37	37	188	337
2	Terminal	--	--	-	-	411	0	411	0.0	1.0	37	37	0	337
3	Terminal	--	--	-	-	188	188	0	1.0	0.0	0	0	188	0

Decision Tree for Ice Clouds
0.1% Multiplicative Noise
Table 5.11

NODE	DISCRIMINATING FEATURE	CUT VALUE	MAXIMUM KS DISTANCE	NEXT NODE IF LESS THAN OR EQUAL	NEXT NODE IF GREATER THAN	TOTAL POPULATION AT NODE	MEMBERS IN CLASS	MEMBERS NOT IN CLASS	P_1	P_2	CLEAR COLUMNS OVER LOAM	CLEAR COLUMNS OVER SNOW	WATER CLOUDS	ICE CLOUDS
1	$I_{1.6}/I_{2.1}$	3.02	0.953	3	2	599	337	262	1.0	1.0	37	37	188	337
2	$T_{10.5}/T_{11.5}$	9.98×10^{-1}	0.747	5	4	278	16	262	0.05	1.0	37	37	188	16
3	Terminal	--	--	-	-	321	321	0	0.95	0.0	0	0	0	321
4	$I_{1.6}/I_{2.1}$	3.52	0.800	7	6	65	15	50	0.04	0.19	27	23	0	15
5	Terminal	--	--	-	-	213	1	212	0.0	0.81	10	14	188	1
6	$I_{1.0}/T_{12.5}$		1.00	9	8	53	3	50	0.01	0.19	27	23	0	3
7	Terminal	--	--	-	-	12	12	0	0.04	0.0	0	0	0	12
8	Terminal	--	--	-	-	3	3	0	0.01	0.0	0	0	0	3
9	Terminal	--	--	-	-	50	0	50	0.0	0.19	27	23	0	0

Decision Tree for Clear Columns Over Loam

0.1% Multiplicative Noise

Table 5.12

NODE	DISCRIMINATING FEATURE	CUT VALUE	MAXIMUM KS DISTANCE	NEXT NODE IF LESS THAN OR EQUAL	NEXT NODE IF GREATER THAN	TOTAL POPULATION AT NODE	MEMBERS IN CLASS	MEMBERS NOT IN CLASS	P_1	P_2	CLEAR COLUMNS OVER LOAM	CLEAR COLUMNS OVER SNOW	WATER CLOUDS	ICE CLOUDS
1	$I_{1.0}/I_{12.5}$	5.70×10^5	1.00	3	2	599	37	562	1.0	1.0	37	37	188	337
2	Terminal	--	--	-	-	562	0	562	0.0	1.0	0	37	188	337
3	Terminal	--	--	-	-	37	37	0	1.0	0.0	37	0	0	0

Decision Tree for Clear Columns Over Snow

0.1% Multiplicative Noise

Table 5.13

NODE	DISCRIMINATING FEATURE	CUT VALUE	MAXIMUM KS DISTANCE	NEXT NODE IF LESS THAN OR EQUAL	NEXT NODE IF GREATER THAN	TOTAL POPULATION AT NODE	MEMBERS IN CLASS	MEMBERS NOT IN CLASS	P_1	P_2	CLEAR COLUMNS OVER LOAM	CLEAR COLUMNS OVER SNOW	WATER CLOUDS	ICE CLOUDS
1	$I_{1.0}/I_{2.1}$	1.43×10^2	1.00	3	2	599	37	562	1.0	1.0	37	37	188	337
2	Terminal	--	--	-	-	37	37	0	1.0	0.0	0	37	0	0
3	Terminal	--	--	-	-	562	0	562	0.0	1.0	37	0	188	337

Decision Tree for Water Clouds

0.5% Multiplicative Noise

Table 5.14

NODE	DISCRIMINATING FEATURE	CUT VALUE	MAXIMUM KS DISTANCE	NEXT NODE IF LESS THAN OR EQUAL	NEXT NODE IF GREATER THAN	TOTAL POPULATION AT NODE	MEMBERS IN CLASS	MEMBERS NOT IN CLASS	P_1	P_2	CLEAR COLUMNS OVER LOAM	CLEAR COLUMNS OVER SNOW	WATER CLOUDS	ICE CLOUDS
1	$I_{1.6}/T_{12.5}$	1.31×10^6	1.00	2	3	599	188	411	1.0	1.0	37	37	188	337
2	Terminal	--	--	-	-	411	0	411	0	1.0	37	37	0	337
3	Terminal	--	--	-	-	188	188	0	1.0	0.0	0	0	188	0

Decision Tree for Ice Clouds

0.5% Multiplicative Noise

Table 5.15

NODE	DISCRIMINATING FEATURE	CUT VALUE	MAXIMUM KS DISTANCE	NEXT NODE IF LESS THAN OR EQUAL	NEXT NODE IF GREATER THAN	TOTAL POPULATION AT NODE	MEMBERS IN CLASS	MEMBERS NOT IN CLASS	P_1	P_2	CLEAR COLUMNS OVER LOAM	CLEAR COLUMNS OVER SNOW	WATER CLOUDS	ICE CLOUDS
1	$I_{1.6}/I_{2.1}$	3.01	0.953	3	2	599	337	262	1.0	1.0	37	37	188	337
2	$T_{3.8}/T_{12.5}$	5.18×10^3	0.733	5	4	278	16	262	0.05	1.0	37	37	188	16
3	Terminal	--	--	-	-	321	321	0	0.95	0.0	0	0	0	321
4	Terminal	--	--	-	-	192	0	192	0.0	0.73	0	4	188	0
5	$I_{1.6}/I_{2.1}$	3.50	0.813	7	6	86	16	70	0.0	0.27	37	33	0	16
6	$I_{1.0}/T_{12.5}$	1.35×10^6	1.00	9	8	73	3	70	0.01	0.27	37	33	0	3
7	Terminal	--	--	-	-	13	13	0	0.04	0.0	0	0	0	13
8	Terminal	--	--	-	-	3	3	0	0.01	0.0	0	0	0	3
9	Terminal	--	--	-	-	70	0	70	0.0	0.27	37	33	0	0

Decision Tree for Clear Columns Over Loam

0.5% Multiplicative Noise

Table 5.16

NODE	DISCRIMINATING FEATURE	CUT VALUE	MAXIMUM KS DISTANCE	NEXT NODE IF LESS THAN OR EQUAL	NEXT NODE IF GREATER THAN	TOTAL POPULATION AT NODE	MEMBERS IN CLASS	MEMBERS NOT IN CLASS	P_1	P_2	CLEAR COLUMNS OVER LOAM	CLEAR COLUMNS OVER SNOW	WATER CLOUDS	ICE CLOUDS
1	$I_{1.0}/T_{12.5}$	5.74×10^5	1.00	3	2	599	37	562	1.0	1.0	37	37	188	337
2	Terminal	--	--	-	-	562	0	562	0.0	1.0	0	37	188	337
3	Terminal	--	--	-	-	37	37	0	1.0	0.0	37	0	0	0

Decision Tree for Clear Columns Over Snow

0.5% Multiplicative Noise

Table 5.17

NODE	DISCRIMINATING FEATURE	CUT VALUE	MAXIMUM KS DISTANCE	NEXT NODE IF LESS THAN OR EQUAL	NEXT NODE IF GREATER THAN	TOTAL POPULATION AT NODE	MEMBERS IN CLASS	MEMBERS NOT IN CLASS	P_1	P_2	CLEAR COLUMNS OVER LOAM	CLEAR COLUMNS OVER SNOW	WATER CLOUDS	ICE CLOUDS
1	$I_{1.0}/I_{2.1}$	1.43×10^2	1.00	3	2	599	37	562	1.0	1.0	37	37	188	337
2	Terminal	--	--	-	-	37	0	0	1.0	0.0	0	37	0	0
3	Terminal	--	--	-	-	562	0	562	0.0	1.0	37	0	188	337

Decision Tree for Water Clouds

1.0% Multiplicative Noise

Table 5.18

NODE	DISCRIMINATING FEATURE	CUT VALUE	MAXIMUM KS DISTANCE	NEXT NODE IF LESS THAN OR EQUAL	NEXT NODE IF GREATER THAN	TOTAL POPULATION AT NODE	MEMBERS IN CLASS	MEMBERS NOT IN CLASS	P_1	P_2	CLEAR COLUMNS OVER LOAM	CLEAR COLUMNS OVER SNOW	WATER CLOUDS	ICE CLOUDS
1	$I_{1.6}/T_{12.5}$	1.32×10^6	1.00	2	3	599	188	411	1.0	1.0	37	37	188	337
2	Terminal	--	--	-	-	411	0	411	0.0	1.0	37	37	0	337
3	Terminal	--	--	-	-	188	188	0	1.0	0.0	0	0	188	0

Decision Tree for Ice Clouds

1.0% Multiplicative Noise

Table 5.19

NODE	DISCRIMINATING FEATURE	CUT VALUE	MAXIMUM KS DISTANCE	NEXT NODE IF LESS THAN OR EQUAL	NEXT NODE IF GREATER THAN	TOTAL POPULATION AT NODE	MEMBERS IN CLASS	MEMBERS NOT IN CLASS	P_1	P_2	CLEAR COLUMNS OVER LOAM	CLEAR COLUMNS OVER SNOW	WATER CLOUDS	ICE CLOUDS
1	$I_{1.6}/I_{2.1}$	3.01	0.958	3	2	599	337	262	1.0	1.0	37	37	188	337
2	$I_{3.8}/T_{11.5}$	5.12×10^3	0.731	5	4	277	17	260	.05	0.99	37	37	186	17
3	$I_{3.8}/T_{12.5}$	3.42×10^4	1.00	7	6	322	320	2	.95	0.01	0	0	2	320
4	Terminal	--	--	-	-	190	0	190	0.0	0.73	0	4	186	0
5	$I_{1.6}/I_{2.1}$	3.48	0.781	9	8	87	17	70	0.05	0.27	37	33	0	17
6	Terminal	--	--	-	-	2	0	2	0.0	0.01	0	0	2	0
7	Terminal	--	--	-	-	320	320	0	0.95	0.0	0	0	0	320
8	$I_{1.0}/T_{12.5}$	1.38×10^6	1.00	11	10	70	3	67	0.01	0.26	34	33	0	3
9	$I_{1.0}/T_{6.7}$	5.59×10^5	1.00	13	12	17	14	3	0.04	0.01	3	0	0	14
10	Terminal	--	--	-	-	3	3	0	0.01	0.0	0	0	0	3
11	Terminal	--	--	-	-	67	0	67	0.0	0.26	34	33	0	0
12	Terminal	--	--	-	-	14	14	0	0.04	0.0	0	0	0	14
13	Terminal	--	--	-	-	3	0	3	0.0	0.01	3	0	0	0

Decision Tree for Clear Columns Over Loam

1.0% Multiplicative Noise

Table 5.20

NODE	DISCRIMINATING FEATURE	CUT VALUE	MAXIMUM KS DISTANCE	NEXT NODE IF LESS THAN OR EQUAL	NEXT NODE IF GREATER THAN	TOTAL POPULATION AT NODE	MEMBERS IN CLASS	MEMBERS NOT IN CLASS	P_1	P_2	CLEAR COLUMNS OVER LOAM	CLEAR COLUMNS OVER SNOW	WATER CLOUDS	ICE CLOUDS
1	$I_{1.0}/T_{6.7}$	5.71×10^5	1.00	3	2	599	37	562	1.0	1.0	37	37	188	337
2	Terminal	--	--	-	-	562	0	562	0.0	1.0	0	37	188	337
3	Terminal	--	--	-	-	37	37	0	1.0	0.0	37	0	0	0

Decision Tree for Clear Columns Over Snow

1.0% Multiplicative Noise

Table 5.21

NODE	DISCRIMINATING FEATURE	CUT VALUE	MAXIMUM KS DISTANCE	NEXT NODE IF LESS THAN OR EQUAL	NEXT NODE IF GREATER THAN	TOTAL POPULATION AT NODE	MEMBERS IN CLASS	MEMBERS NOT IN CLASS	P_1	P_2	CLEAR COLUMNS OVER LOAM	CLEAR COLUMNS OVER SNOW	WATER CLOUDS	ICE CLOUDS
1	$I_{1.0}/I_{2.1}$	1.43×10^2	1.00	3	2	599	37	562	1.0	1.0	37	37	188	337
2	Terminal	--	--	-	-	37	37	0	1.0	0.0	0	37	0	0
3	Terminal	--	--	-	-	562	0	562	0.0	1.0	37	0	188	337

Decision Tree for Water Clouds

5.0 Multiplicative Noise

Table 5.22

NODE	DISCRIMINATING FEATURE	CUT VALUE	MAXIMUM KS DISTANCE	NEXT NODE IF LESS THAN OR EQUAL	NEXT NODE IF GREATER THAN	TOTAL POPULATION AT NODE	MEMBERS IN CLASS	MEMBERS NOT IN CLASS	P_1	P_2	CLEAR COLUMNS OVER LOAM	CLEAR COLUMNS OVER SNOW	WATER CLOUDS	ICE CLOUDS
1	$I_{1.6}$	1.30×10^6	1.00	2	3	599	188	411	1.0	1.0	37	37	188	337
2	Terminal	--	--	-	-	411	0	411	0.0	1.0	37	37	0	337
3	Terminal	--	--	-	-	188	188	0	1.0	0.0	0	0	188	0

Decision Tree for Ice Clouds

5.0% Multiplicative Noise

Table 5.23

NODE	DISCRIMINATING FEATURE	CUT VALUE	MAXIMUM KS DISTANCE	NEXT NODE IF LESS THAN OR EQUAL	NEXT NODE IF GREATER THAN	TOTAL POPULATION AT NODE	MEMBERS IN CLASS	MEMBERS NOT IN CLASS	P_1	P_2	CLEAR COLUMNS OVER LOAM	CLEAR COLUMNS OVER SNOW	WATER CLOUDS	ICE CLOUDS
1	$I_{1.6}/I_{2.1}$	3.13	0.895	3	2	599	337	262	1.0	1.0	37	37	188	337
2	$I_{1.0}/I_{1.6}$	5.34	0.713	5	4	263	16	247	0.05	0.94	34	37	176	16
3	$I_{1.6}/I_{2.1}$	2.85	0.809	7	6	336	321	15	0.95	0.06	3	0	12	321
4	$I_{1.6}/I_{2.1}$	3.42	0.714	9	8	87	16	71	0.05	0.27	34	37	0	16
5	Terminal	--	--	-	-	176	0	176	0.0	0.67	0	0	176	0
6	$I_{3.8}/T_{11.5}$	3.47×10^4	0.786	11	10	54	40	14	0.12	0.05	3	0	11	40
7	Terminal	--	--	-	-	282	281	1	0.84	0.0	0	0	1	281
8	$I_{1.0}/T_{6.7}$	1.42×10^6	0.984	13	12	67	3	64	0.01	0.24	27	37	0	3
9	$I_{1.0}$	1.53×10^8	1.00	15	14	20	13	7	0.04	0.03	7	0	0	13
10	Terminal	--	--	-	-	11	0	11	0.0	0.04	0	0	11	0
11	$I_{1.0}/T_{10.5}$	5.93×10^5	1.00	17	16	43	40	3	0.12	0.01	3	0	0	40

Decision Tree for Ice Clouds

5.0% Multiplicative Noise

Table 5.23

(Continued)

NODE	DISCRIMINATING FEATURE	CUT VALUE	MAXIMUM KS DISTANCE	NEXT NODE IF LESS THAN OR EQUAL	NEXT NODE IF GREATER THAN	TOTAL POPULATION AT NODE	MEMBERS IN CLASS	MEMBERS NOT IN CLASS	P_1	P_2	CLEAR COLUMNS OVER LOAM	CLEAR COLUMNS OVER SNOW	WATER CLOUDS	ICE CLOUDS
12	Terminal	--	--	-	-	4	3	1	0.01	0.0	0	1	0	3
13	Terminal	--	--	-	-	63	0	63	0.0	0.24	27	36	0	0
14	Terminal	--	--	-	-	13	13	0	0.04	0.0	0	0	0	13
15	Terminal	--	--	-	-	7	0	7	0.0	0.03	7	0	0	0
16	Terminal	--	--	-	-	40	40	0	0.12	0.0	0	0	0	40
17	Terminal	--	--	-	-	3	0	3	0.0	0.01	3	0	0	0

Decision Tree for Clear Columns Over Loam

5.0% Multiplicative Noise

Table 5.24

NODE	DISCRIMINATING FEATURE	CUT VALUE	MAXIMUM KS DISTANCE	NEXT NODE IF LESS THAN OR EQUAL	NEXT NODE IF GREATER THAN	TOTAL POPULATION AT NODE	MEMBERS IN CLASS	MEMBERS NOT IN CLASS	P_1	P_2	CLEAR COLUMNS OVER LOAM	CLEAR COLUMNS OVER SNOW	WATER CLOUDS	ICE CLOUDS
1	$I_{.72}$	3.41×10^8	.991	3	2	599	37	562	1.0	1.0	37	37	188	337
2	Terminal	--	--	-	-	557	0	557	0.0	0.99	0	37	188	332
3	$I_{.55}/I_{2.1}$	1.37×10^2	1.00	5	4	42	37	5	1.0	0.01	37	0	0	5
4	Terminal	--	--	-	-	37	37	0	1.0	0.0	37	0	0	0
5	Terminal	--	--	-	-	5	0	5	0.0	0.01	0	0	0	5

Decision Tree for Clear Columns Over Snow

5.0% Multiplicative Noise

Table 5.25

NODE	DISCRIMINATING FEATURE	CUT VALUE	MAXIMUM KS DISTANCE	NEXT NODE IF LESS THAN OR EQUAL	NEXT NODE IF GREATER THAN	TOTAL POPULATION AT NODE	MEMBERS IN CLASS	MEMBERS NOT IN CLASS	P_1	P_2	CLEAR COLUMNS OVER LOAM	CLEAR COLUMNS OVER SNOW	WATER CLOUDS	ICE CLOUDS
1	$I_{2.1}/T_{6.7}$	9.73×10^3	.998	3	2	599	37	562	1.0	1.0	37	37	188	337
2	Terminal	--	--	-	-	37	37	1	1.0	0.0	0	37	0	1
3	Terminal	--	--	-	-	561	0	561	0.0	1.0	37	0	188	336

Decision Tree for Water Clouds

10.0% Multiplicative Noise

Table 5.26

NODE	DISCRIMINATING FEATURE	CUT VALUE	MAXIMUM KS DISTANCE	NEXT NODE IF LESS THAN OR EQUAL	NEXT NODE IF GREATER THAN	TOTAL POPULATION AT NODE	MEMBERS IN CLASS	MEMBERS NOT IN CLASS	P_1	P_2	CLEAR COLUMNS OVER LOAM	CLEAR COLUMNS OVER SNOW	WATER CLOUDS	ICE CLOUDS
1	$I_{1.6}$	3.31×10^8	0.992	3	2	599	188	411	1.0	1.0	37	37	188	337
2	Terminal	--	--	-	-	188	187	1	0.99	0.0	0	0	187	1
3	Terminal	--	--	-	-	411	1	410	0.01	0.99	37	37	1	336

Decision Tree for Ice Clouds
10.0% Multiplicative Noise
Table 5.27

NODE	DISCRIMINATING FEATURE	CUT VALUE	MAXIMUM KS DISTANCE	NEXT NODE IF LESS THAN OR EQUAL	NEXT NODE IF GREATER THAN	TOTAL POPULATION AT NODE	MEMBERS IN CLASS	MEMBERS NOT IN CLASS	P_1	P_2	CLEAR COLUMNS OVER LOAM	CLEAR COLUMNS OVER SNOW	WATER CLOUDS	ICE CLOUDS
1	$I_{1.6}/I_{2.1}$	3.08	0.725	3	2	599	337	262	1.0	1.0	37	37	188	337
2	$I_{1.6}$	3.31×10^8	0.693	5	4	270	45	225	0.13	0.86	32	37	156	45
3	$I_{1.6}$	2.82×10^8	0.844	7	6	329	292	37	0.87	0.14	5	0	32	292
4	Terminal	--	--	-	-	156	0	156	0.0	0.60	0	0	156	0
5	$I_{2.1}$	4.69×10^6	0.800	9	8	114	45	69	0.13	0.26	32	37	0	45
6	$I_{2.1}$	1.12×10^8	0.969	11	10	38	6	32	0.02	0.12	0	0	32	6
7	$I_{.22}/T_{10.5}$	1.13×10^6	0.993	13	12	291	286	5	0.85	0.02	5	0	0	286
8	Terminal	--	--	-	-	36	36	0	0.11	0.0	0	0	0	36
9	$I_{1.0}/I_{3.8}$	3.76×10^2	0.899	15	14	78	9	69	0.03	0.26	32	37	0	9
10	Terminal	--	--	-	-	31	0	31	0.0	0.0	0	0	31	0
11	Terminal	--	--	-	-	7	6	1	0.02	0.0	0	0	1	6

Decision Tree for Ice Clouds

10.0% Multiplicative Noise

Table 5.27

(Continued)

NODE	DISCRIMINATING FEATURE	CUT VALUE	MAXIMUM KS DISTANCE	NEXT NODE IF LESS THAN OR EQUAL	NEXT NODE IF GREATER THAN	TOTAL POPULATION AT NODE	MEMBERS IN CLASS	MEMBERS NOT IN CLASS	P_1	P_2	CLEAR COLUMNS OVER LOAM	CLEAR COLUMNS OVER SNOW	WATER CLOUDS	ICE CLOUDS
12	Terminal	--	--	-	-	284	284	0	0.84	0.0	0	0	0	284
13	Terminal	--	--	-	-	7	2	5	0.01	0.02	5	0	0	2
14	$I_{2.1}/T_{6.7}$	9.42×10^3	1.00	17	16	16	9	7	0.03	0.03	0	7	0	9
15	Terminal	--	--	-	-	62	0	62	0.0	0.24	32	30	0	0
16	Terminal	--	--	-	-	9	9	0	0.03	0.0	0	0	0	9
17	Terminal	--	--	-	-	7	0	7	0.0	0.03	0	7	0	0

Decision Tree for Clear Columns Over Loam

10.0% Multiplicative Noise

Table 5.28

NODE	DISCRIMINATING FEATURE	CUT VALUE	MAXIMUM KS DISTANCE	NEXT NODE IF LESS THAN OR EQUAL	NEXT NODE IF GREATER THAN	TOTAL POPULATION AT NODE	MEMBERS IN CLASS	MEMBERS NOT IN CLASS	P_1	P_2	CLEAR COLUMNS OVER LOAM	CLEAR COLUMNS OVER SNOW	WATER CLOUDS	ICE CLOUDS
1	$I_{1.0}$	1.71×10^8	0.963	3	2	599	37	562	1.0	1.0	37	37	188	337
2	Terminal	--	--	-	-	541	0	541	0.0	0.96	0	37	188	316
3	$I_{2.1}/T_{6.7}$	2.21×10^4	0.952	5	4	58	37	21	1.0	0.04	37	0	0	21
4	Terminal	--	--	-	-	20	0	20	0.0	0.04	0	0	0	20
5	Terminal	--	--	-	-	38	37	1	1.0	0.0	37	0	0	1

Decision Tree for Clear Columns Over Snow

10.0% Multiplicative Noise

Table 5.29

NODE	DISCRIMINATING FEATURE	CUT VALUE	MAXIMUM KS DISTANCE	NEXT NODE IF LESS THAN OR EQUAL	NEXT NODE IF GREATER THAN	TOTAL POPULATION AT NODE	MEMBERS IN CLASS	MEMBERS NOT IN CLASS	P_1	P_2	CLEAR COLUMNS OVER LOAM	CLEAR COLUMNS OVER SNOW	WATER CLOUDS	ICE CLOUDS
1	$I_{1.0}/I_{2.1}$	1.24×10^2	0.996	3	2	599	37	562	1.0	1.0	37	37	188	337
2	$I_{1.6}/T_{6.7}$	3.80×10^4	1.00	5	4	39	37	2	1.0	0.0	0	37	0	2
3	Terminal	--	--	-	-	560	0	560	0.0	1.00	37	0	188	335
4	Terminal	--	--	-	-	2	0	2	0.0	0.0	0	0	0	2
5	Terminal	--	--	-	-	37	37	0	1.0	0.0	0	37	0	0

As discussed in the following section, the features selected represent those with real discrimination ability, rather than those whose discrimination ability comes from small differences in boundary assumptions, etc. A number of quantities are depicted in these figures. The starting node of the tree is denoted by an S and the terminal node of the tree by a T. The quantity in the box is the partitioning feature with its cut value. For the first experiment, the bracketed quantities below a node are the number of clear columns over loam, clear columns over snow, water clouds and ice clouds respectively that descended to that node. The probability P_1 given at a node is the conditional probability that an observation of the class to be discriminated will appear at the node given that the tree was entered with a member of that class. The probability P_2 is the conditional probability of the other classes reaching that node.

Tables 5-6 through 5-29 give the decision trees for the first experiment in tabular form. These are given for noises of 0, 0.1, 0.5, 1.0, 5.0 and 10.0 percent.

5.4 Discussion of the 4 Classes Discrimination

The decision trees for almost all cases are quite simple, consisting usually of only two or three levels. The tree for water clouds are universally the shortest. For no noise, the feature used for discrimination is $T_{10.5}/T_{12.5}$. For noise values from 0.1% through 1.0%, the choice shifts to $I_{1.6}/T_{12.5}$. At 5% and 10% noise, the feature used is $I_{1.6}$ alone.

The $T_{10.5}/T_{12.5}$ feature is probably an artifact resulting from an assumption of 100% emissivity of the surface. The feature selected results from water clouds being both relatively bright at 1.6 μ and cold in the thermal region, thus yielding a high ratio and high $I_{1.6}$ value alone.

The decision trees for ice clouds are somewhat more complicated. The primary feature used is the ratio of the two near-IR channels 1.6 μ and 2.1 μ. The few cases not discriminated by this test, are resolved by tests using ratios of a visible or near-IR channel to a thermal one. The primary feature results from the fact that the cloud model used calculated a lower reflectance for ice clouds at 1.6 μ relative to their 2.1 μ reflectance than was calculated for water clouds or assumed for loam and snow.

The primary discriminant for clear columns over loam is the ratio of $I_{1.0}$ to a thermal channel (usually $T_{12.5}$). At noises at or above 5.0%, this changes to the value of a visual channel alone. This comes from the fact that the loam surface is both visually dark and thermally warm.

The final case of clear columns over snow also uses a $I_{1.6}/I_{2.1}$ ratio for a low noise discriminating feature. The lack of robustness of this feature is evidenced in the cases of increased noise by a switch to the ratio of $I_{.55}$ or $I_{1.0}$ to $I_{2.1}$. This feature is quite robust through the 10% noise case and discriminates well. It is due to the sudden drop in the surface reflectivity of snow at about 1.6 μ. This feature has been noted previously by Valovcin (1976) and Hunt et al. (1974).

VI. INVERSIONS USING COMBINED MICROWAVE, VISIBLE AND INFRARED DATA

6.1 Inversion for Continuous-Valued Parameters

A series of inversions were tried using the combination of microwave, visible and infrared channels studied in the preceding chapters. The data set used was the one employed for analyzing visible and infrared data (cf., Table 4.2.1), since the primary interest was to assess the potential of all sensors when each sensor may contribute some information.

A series of linear regressions was run for liquid water content, cloud thickness and height and water vapor and the results are given in Tables 6.1.1 - 6.1.4. It can be seen that the 10.7 GHz channel is a significant predictor of liquid water content (R^2 = 99.6), and that the 6.7 micron and 19 GHz channels have some predictive power. The 3.8 micron channel is the most significant for predicting cloud thickness (R^2 = 76.2), but this parameter cannot be determined with great accuracy. Cloud top height was correlated most highly with the 10.5 micron channel intensity (R^2 = 93.5) and to a lesser extent with the 94 GHz brightness temperature and 1.6 micron channel intensity. Water vapor was most highly correlated with the four microwave channel brightness temperatures (R^2 = 88.5).

Next, the effect of adding noise to the test set was evaluated, using the same linear regression models. Noise levels of 0.1% and 1% rms were used, with the results shown in Tables 6.1.5 - 6.1.8. The 0.1% noise is seen to have an insignificant effect in all cases. However, the 1% noise caused significant degradation in the inversions for liquid water content

Table 6.1.1

INVERSION FOR LWC VIA LINEAR
REGRESSION USING VISIBLE, INFRARED
AND MICROWAVE DATA

Order	Variable	Coefficient	R^2	σ
0	mean	*1.017×10^2	0	136.48
1	T_{10}	1.764×10^1	.996	8.63
2	T_{19}	-1.369	.998	6.54
3	$A_{3.8}$	-9.965×10^2	.999	3.98
4	T_{37}	-1.256	.999	3.50
5	$I_{10.5}$	3.639×10^1	.999	3.33
6	$A_{2.1}$	1.959×10^2	.999	2.97
7	$I_{12.5}$	-2.304×10^1	.999	2.92

Intercept Estimate -1.452×10^3

*Mean value is given for reference purposes only.

Table 6.1.2

INVERSION FOR CT VIA LINEAR
REGRESSION USING VISIBLE, INFRARED
AND MICROWAVE DATA

Order	Variable	Coefficient	R^2	σ
0	mean	$*1.131 \times 10^{1}$	0	9.36
1	$A_{3.8}$	-1.154×10^{3}	.762	4.57
2	$I_{12.5}$	3.248×10^{1}	.799	4.21
3	T_{94}	-4.754×10^{-1}	.818	4.02
4	T_{37}	-1.063	.825	3.95
5	T_{19}	1.157	.846	3.73
6	$I_{6.7}$	-2.236×10^{1}	.869	3.44
7	T_{10}	-1.230	.883	3.27
8	$A_{1.0}$	6.157×10^{1}	.891	3.15

Intercept Estimate 4.718×10^{2}

*Mean value is given for reference purposes only.

Table 6.1.3

INVERSION FOR CH VIA LINEAR
REGRESSION USING VISIBLE, INFRARED
AND MICROWAVE DATA

Order	Variable	Coefficient	R^2	σ
0	mean	*3.783×10^1	0	14.26
1	$I_{10.5}$	-1.581×10^1	.935	3.67
2	$A_{1.0}$	-1.592×10^1	.942	3.47

Intercept estimate 1.231×10^2

*Mean value is given for reference purposes only.

Table 6.1.4

INVERSION FOR WV VIA LINEAR
REGRESSION USING VISIBLE, INFRARED
AND MICROWAVE DATA

Order	Variable	Coefficient	R^2	σ
0	mean	*0	0	.419
1	T_{19}	1.988×10^{-1}	.092	.399
2	T_{10}	-3.163×10^{-1}	.583	.271
3	T_{37}	-4.384×10^{-2}	.876	.149
4	T_{94}	6.761×10^{-3}	.885	.143
5	$A_{3.8}$	-4.625	.893	.139

Intercept estimate 1.740×10^{1}

*Mean value is given for reference purposes only.

Table 6.1.5

EFFECT OF MEASUREMENT NOISE ON
INTEGRATED LIQUID WATER REGRESSIONS

- combined measurements
- $\sigma_{model} = 0$

Order	σ		
	$\sigma_n = 0$	$\sigma_n = 0.1\%$	$\sigma_n = 1\%$
0	136.48	136.48	136.48
1	8.63	8.76	17.89
2	6.54	6.87	22.60
3	3.98	4.56	21.08
4	3.50	4.38	21.70
5	3.33	4.00	21.65
6	2.97	3.51	21.47
7	2.92	3.78	23.00

Table 6.1.6

EFFECT OF MEASUREMENT NOISE ON
CLOUD THICKNESS LINEAR REGRESSIONS

- combined measurements
- $\sigma_{model} = 0$

Order	σ		
	$\sigma_n = 0$	$\sigma_n = 0.1\%$	$\sigma_n = 1\%$
0	9.36	9.36	9.36
1	4.57	4.57	4.62
2	4.21	4.22	4.29
3	4.02	4.04	4.23
4	3.95	3.98	4.35
5	3.73	3.85	4.26
6	3.44	3.74	4.40
7	3.27	3.50	5.60
8	3.15	3.23	6.12

Table 6.1.7

EFFECT OF MEASUREMENT NOISE ON
CLOUD HEIGHT LINEAR REGRESSIONS

- combined measurements
- $\sigma_{model} = 0$

Order	σ		
	$\sigma_n = 0$	$\sigma_n = 0.1\%$	$\sigma_n = 1\%$
0	14.26	14.26	14.26
1	3.67	3.67	3.76
2	3.47	3.48	3.56

Table 6.1.8

EFFECT OF MEASUREMENT NOISE ON
WATER VAPOR LINEAR REGRESSIONS

- combined measurements
- $\sigma_{model} = 0$

Order	σ		
	$\sigma_n = 0$	$\sigma_n = 0.1\%$	$\sigma_n = 1\%$
0	.419	.419	.419
1	.399	.399	.399
2	.271	.273	.433
3	.149	.151	.478
4	.143	.148	.495
5	.139	.147	.474

and water vapor. There was only a small degradation in the cloud thickness inversions, and the inversions for cloud height were essentially unaffected.

6.2 Scenario for Inference of Geophysical Parameters Using Combined Visible, Infrared and Microwave Data

A possible scenario for geophysical parameter inversion is given in Figures 6.2.1 and 6.2.2. In Figure 6.2.1, a method for inversion over the ocean surface is given. The first step is separation of clouds from clear columns, with further classification of cloud type into either water cloud or ice cloud. This may be done using the decision tree approach given in Section V. Following this step, three conditional inverters are utilized. The water cloud inverter is perhaps the most complex of the three; for this reason, it has received the most extensive treatment in this study (cf., Chapters III, IV). The possibility of inverting for cloud top height and thickness, cloud water content, surface wind speed, atmospheric water vapor and rainfall rate has been studied at length in that Chapter, and the reader is referred to that discussion for further information. The clear column inverter would be used for determination of wind speed and atmospheric water vapor, using all channels, in the same manner as the water cloud inverter. The same approach could be used for the ice cloud inverter, if it was desired to estimate the continuous-valued parameters in that case.

The situation over land is given in Figure 6.2.2 and is seen to be somewhat more complex. The first step is a classification of both cloud and surface type. Here, a total of four categories is used:

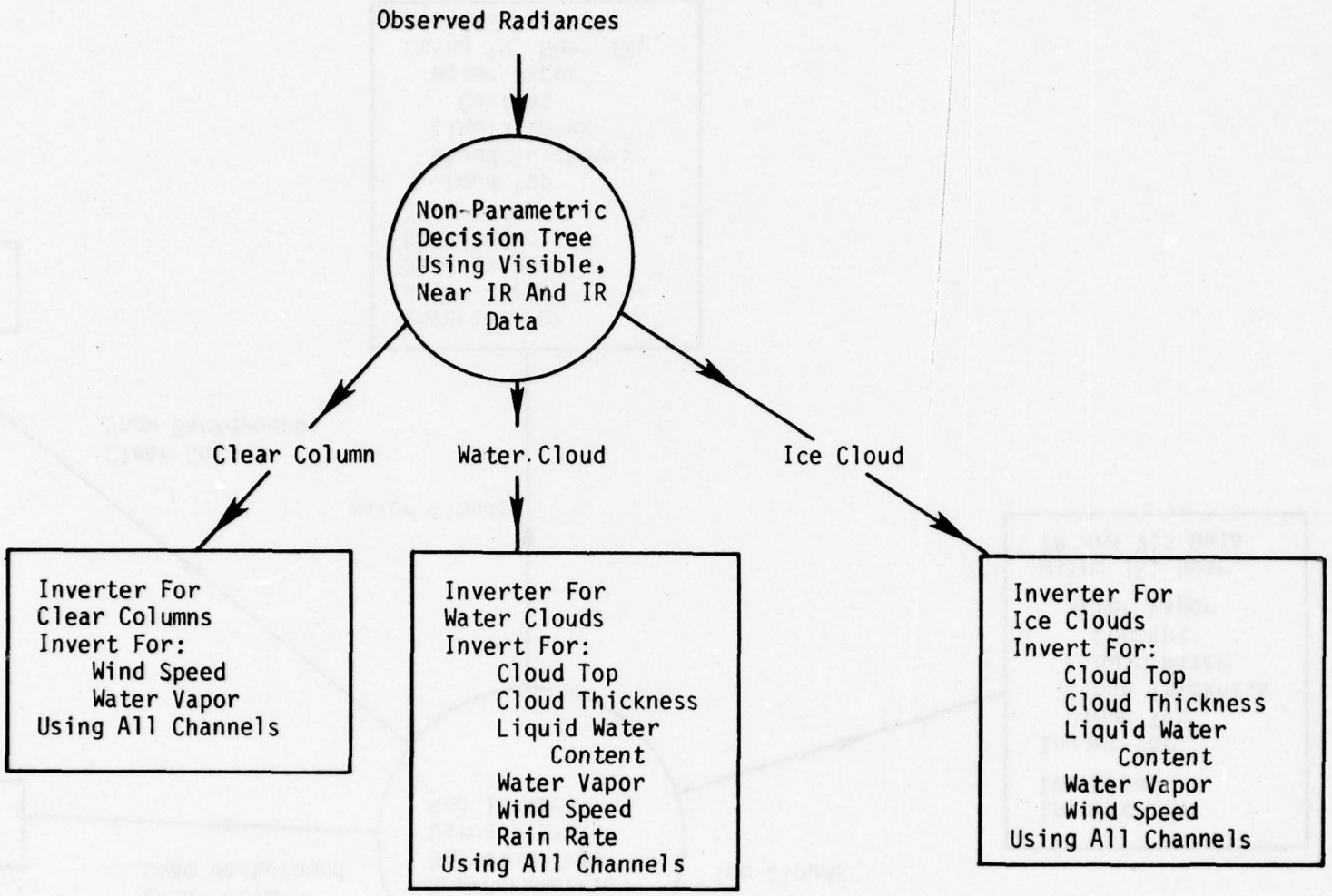


Figure 6.2.1 Inversion Scenario for Inversion Over Ocean

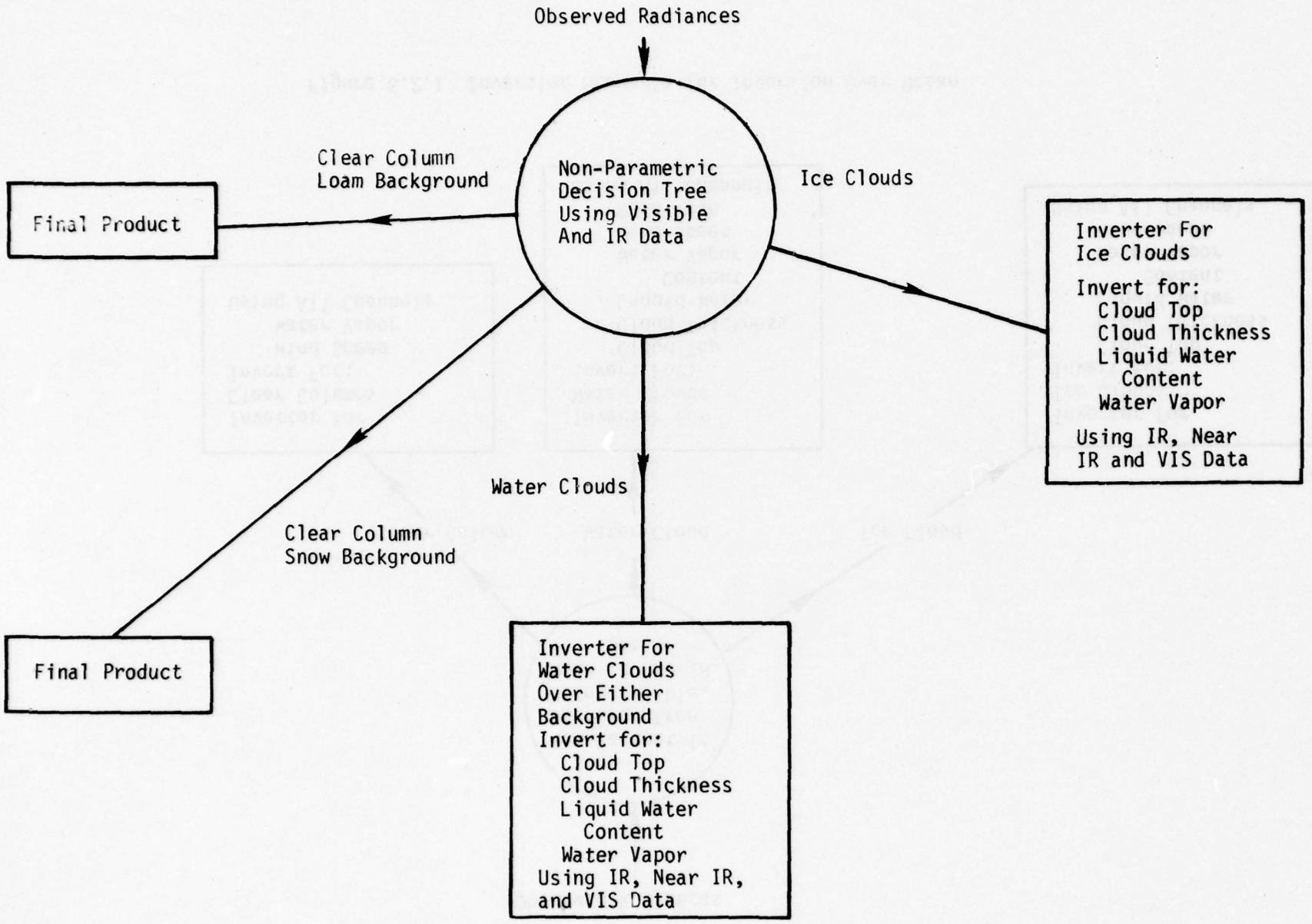


Figure 6.2.2 Inversion Scenario for Inversion Over Land

	Clouds	Surface
1.	none	loam
2.	none	snow
3.	water	----
4.	ice	----

The classification may be accomplished using the decision tree approach of Chapter V. Extensive results are given there to illustrate the structure and error rates of such trees. Once this classification has been made, the next step is to invert for the continuous-valued parameters using one of the four conditional inverters shown in the figure. These inverters may be designed using the methods of Chapter II.

VII. PLANS FOR TESTING AND VERIFICATION

The inversion and classification schemes developed in this study were constructed through the use of simulated observations. For this reason, the results obtained through the use of the methodologies are in some sense hypothetical. Although every effort was made to provide realistic simulations, there are factors such as the spectral response of the optical filters of an instrument which has neither been built nor even specified, that are impossible or expensive to simulate. At such time as observations from a single package are available, however, it will be advisable to implement and test the algorithms to determine their accuracy.

Two major classes of algorithms were presented in this study. The first class provided estimates of continuous valued variables such as wind speed, cloud height, etc. The second provides a means of scene classification. One common characteristic of both is that they are both area averages.

In order to implement the algorithms suggested by this study, several steps will be necessary:

1. Specification of the actual instrument parameters.

2. Creation of new forward models for those frequencies/wavelengths or look angles not used in this study but present on the instrument.

3. Creation of new regression estimators to account for any different selection of frequencies/wavelengths or look angles in the instrument from those considered in this study.

4. The creation and implementation of a data handling package that

implements the algorithms.

5. The construction of a verification dataset that is matched to the satellite observations in space/time.

6. Fine tuning of forward models and Friedman trees from matched data.

7. Verification and error analysis of results.

Step 1 will not be addressed by this section. Suggestions have been made in the course of this study as to the suitability of various frequencies/wavelengths for the inference of various parameters. The ultimate choice is often modulated by hardware considerations such as local oscillator frequencies, antenna and telescope size and number, etc. It is beyond the scope of this study to address these problems.

Steps 2 and 3 will be required if new or different frequencies are used in the ultimate instrument or if the algorithm is to be tested in other than a nadir looking mode. The methods for performing steps 2 and 3 are given in Sections III and IV of this study. Briefly, they will entail the generation of a data base of simulations and the performance of a stepwise regression analysis using terms of the form identified in this study.

Step 4 is again instrument dependent and will not be considered in depth. The inversion algorithms developed in this study should not present any difficulties in this phase.

Steps 5, 6 and 7 are similar in nature and will be considered here in conjunction with each other. Step 5 represents the collection of data to drive steps 6 and 7. Step 6 represents the fine tuning of the

models to reality. The importance of this step cannot be overemphasized. The radiative transfer models used in the model development, while reasonable in terms of the time and resources available for this study, do not handle the entire spectrum of naturally occurring clouds and surfaces. Further, the instrument itself will introduce effects such as scan angle dependent gains and biases that must be removed. Thus, it is important to produce a set of observations matched with known meteorological conditions. The expected radiance based on the radiative transfer model using the known conditions may be compared with the observed radiance and systematic errors removed. Depending on the error source, these comparisons may also be used to refine absorption coefficient models. Finally, the noise performance of the total system may be evaluated.

Once the calibration of step 6 has been performed, the actual performance analysis of step 7 may begin and the error performance achieved may be compared with the error predicted error performance. Standard error analysis that produces error means and covariances will be sufficient, although it will be of interest to see if there exist special situations that degrade the performance more than others.

Since the data set of step 5 is instrumental in both steps 6 and 7, it will now be considered at greater length. It is obviously desirable that the data set consist of a large collection of "verified" data. It is important, however, to determine the source and accuracy of this data. If the data is collected in a biased or inaccurate form, the calibration of step 6 and error analysis of step 7 will be likewise biased or inaccurate.

The most fundamental requirement of the data set is that it should

contain actual observations of the state of the observed scene that are coincident with the instrument overflight in both space and time. This is a non-trivial requirement. The rapidly varying nature of cloud parameters in both space and time suggest that to be useful, the "coincident" observation must be within an hour in time and ≈30 km in space to be useful. During periods of rapid change such as occur during the passage of a front, even these limits may be too large. This data requirement will imply that the verification data will of necessity be collected independently of normal synoptic observations. In particular, gridded analysis data will be of little use as it will be too "smooth" to represent the actual conditions.

Turning to the exact nature of the observations required, it is possible to identify two general classes. For the purposes of forward model calibration and error analysis in the continuous parameter scheme, a complete specification of the atmospheric state is required. For the purposes of the verification of the scene classification algorithm, the correct scene class need only be known.

Of these two requirements, the scene classification verification data is the easiest to acquire. The requirement for this data is to classify the scene into one of four categories:

1. No Cloud, No Snow
2. No Cloud, Snow
3. Water Cloud
4. Ice Cloud

An overflight of an airplane in conjunction with an instrument orbit will yield a large amount of data. The parameters that need to be classified may be recorded by simple photographic techniques and classified by simple human subjective judgement.

The requirements for verification data for tuning and error analysis of the continuous variables are very difficult to meet. First, a large number of variables must be recorded. These variables are:

1. Temperature profiles
2. Water vapor profiles
3. Surface wind speed
4. Surface character and temperature
5. Cloud profiles in terms of density, altitude and phase
6. Instantaneous rainfall rate.

Further, these variables will be required over both land and ocean surfaces. The requirement for ship station data for ocean surface observations will present (and has historically presented) a large problem in the implementation of the data collection due to their low number. In either case, the data collection will have to be done asynoptically to coincide with the instrument passage.

The present means of data collection for the atmospheric state measurements and their approximate present accuracies are given in Table 7.1. As may be seen, the gas parameter measurements such as temperature are usually of higher accuracy than the particle measurements such as the integrated liquid water content of a cloud. Unfortunately, the importance of the measurements in terms of the verification is almost

in reverse order of their accuracy since the satellite measurements are intended to infer particle parameters. It is an open question as to the adequacy of these present measurement techniques to provide a sufficient data set at reasonable cost. Hopefully, advances in the state of the art in lidar, acoustic radar and in situ aircraft measurements will allow verification on a more accurate and less costly basis.

In summary, a plan for implementation and verification has been outlined. Seven basic steps or phases have been outlined. Of these steps, the most difficult and expensive task will be the collection of a sufficiently large and accurate data set for the tuning and verification of continuous parameters. This difficulty is caused by both the general lack of ship stations and the difficulty in the measurements of particle parameters. Some advance in the state of the art in particle parameter measurements seems desirable.

TABLE 7.1

Parameter	Primary Measurement Instrumentation	Estimated Accuracy
Temperature Profile	Radiosonde borne thermistor	1°K
Humidty Profile	Radiosonde borne hydristor	±5-10% of true humidity
Average Surface Wind Speed	Cup anemometer	1-2 m/sec depending on variability
Cloud Base	In situ aircraft radar lidar acoustic radar human estimate	50m-500m depending on thickness, altitude and variability within cloud
Cloud Top	In situ aircraft radar lidar acoustic radar	50m-500m dpending on thickness, altitude and variability within cloud
Rainfall Rate	In situ aircraft radar lidar rain gauge	factor of 2 of true value
Integrated Liquid Water within Cloud	In situ aircraft radar lidar	factor of 2 of true value
Surface Temperature	Thermistor	1°K depending on surface character

VIII. CONCLUSIONS AND RECOMMENDATIONS

The numerical experiments conducted during this study for estimation of continuous-valued parameters have produced several significant results, predicated on the assumed data base. These are summarized in Table 8.1 and in the following:

(1) Integrated liquid water content within clouds appears resolvable using microwave data to within about 0.02 gm/cm^2 with no rainfall. With rainfall, the error increases to about 0.1 gm/cm^2. The 1.6 and 3.8 micron channels can also be used in the absence of microwave data.

(2) Ocean surface wind speed can be estimated to within about 3 m/sec using the 19 and 37 GHz channels.

(3) Systematic atmospheric temperature deviations from an assumed profile cannot be estimated to any significant degree from microwave data.

(4) Single-layer cloud thickness can be determined to within about 600 meters from microwave data. Principal discriminator is the 37 GHz channel, with smaller contributions from the 19 and 94 GHz channels. The 12.5 and 3.8 micron channels are also significant predictors, both with and without coincident microwave data.

(5) Cloud top height can be determined to within about 900 meters using the given microwave channels. The most significant predictor is the 94 GHz channel; however, the 19 and 37 GHz channels also contribute. The 10.5 micron channel contributes significantly to cloud height determination, and a contribution is also made from

TABLE 8.1

Summary of Predictors for

Continuous-Valued Parameters

Channel	LWC	WST	DLT	CT	CH	WV	R
.55 μ							
.725							
1.0					X		
1.6	X						
2.1							
3.8	X			X			
6.7							
10.5					XX		
11.5					X		
12.5				X	X		
10.7 GHz	X	X				X	X
19	XX	XX		X		X	XX
37	XX	X		XX	X	X	
94	X			X	X		

XX significant

X moderate

the 1.0 micron channel. When visible, infrared and microwave channels are used in combination, the 10.5 and 1.0 micron channels are significant predictors.

(6) Atmospheric water vapor cannot be estimated to any great degree of accuracy from the given microwave channels. The error standard deviation could be reduced to only about 0.3 g/cm^2 during the experiments conducted in this study. The visible and infrared channels do not contribute.

(7) Rainfall rate can be determined to within about 3 mm/hr. The principal predictor is the 19 GHz channel, with contributions also made by the 10.7 GHz channel.

The discrete-valued parameter section of this report has produced a number of results. The first and most important is that it is feasible to separate clear columns from water clouds and from even thin or low ice clouds. This is important both in terms of performing the conditional inversion schemes of earlier sections and in determining basic desired meteorological parameters.

A second important result of this study has been the background classification algorithms that have been developed. These experiments show that it is possible to discriminate clear columns over a loam background and clear columns over a snow background from each other and various cloud conditions with a very low error rate.

A final result of the discrete values parameter portion of this effort was that no single channel or spectral interval was of overriding importance. All intervals specified were seen to participate at some point in the decision trees formulated.

In addition to the numerical results, several detailed comparisons of methodologies have been performed. The overall result is that the Bayesian methods are preferred over regression techniques due to their desirable robust properties in the presence of noise and modeling errors. Much more work needs to be performed to develop the most efficient techniques and to explore the extension to real-time dynamic parameter tracking. However, it is felt that the results of this study demonstrate that Bayesian techniques should be seriously considered for implementation into the 3DNEPH program to extend its capability for inference of meteorological parameters.

REFERENCES

1. R. K. Anderson, et al. (1974), "Application of Meteorological Satellite Data in Analysis and Forecasting", ESSA Technical Report, NESC-51.

2. J. C. Barnes and C. J. Bowley (1974), "The Application of ERTS Imagery to Monitoring Arctic Sea Ice", ERT Document 0408-F, February.

3. J. C. Barnes, M. D. Smallwood and J. L. Cogan (1975), "Study to Develop Improved Spacecraft Snow Survey Methods Using Skylab/EREP Data", ERT Document No. 0412-F, May.

4. H. H. Blau, R. P. Espinola and E. C. Reifenstein, III (1966), "Near Infrared Scattering by Sunlit Terrestrial Clouds", Applied Optics, Vol. 5, p. 555, April.

5. H. H. Blau and W. A. Hovis (1971), "Cloud Characteristics from Infrared Measurements", Space Research, Akademic-Verlag, Berlin, p. 731.

6. A. L. Booth (1973A), "Cloud Type Pattern Recognition Using Environmental Satellite Data", Proceedings of the First International Joint Conference on Pattern Recognition.

7. A. L. Booth (1973B), "Objective Cloud Type Classification Using Visual and Infrared Satellite Data", Reprints of the Third Conference on Probability and Statistics in Atmospheric Sciences, Boulder, CO, June.

8. J. T. Bunting and J. H. Conover (1976), "The Use of Satellite Data to Map Excessive Cloud Mass", AFCRL-TR-76-0004, January 6.

9. V. J. Cardone (1969), "Specification of the Wind Field Distribution in the Marine Boundary Layer for Wave Forecasting", Report TR69-Y, Geophysical Science Laboratory, New York University.

10. T. C. Chang, P. Gloersen, T. Schmugge, T. T. Wilheit and H. J. Zwally (1976), "Microwave Emission From Snow And Glacier Ice", J. Glaciology, 16, 74, pp. 23-29.

11. A. R. Coburn (1971), "Improved Three Dimensional Nephanalysis Model", AFGWC Tech. Memo 71-2, June 1.

12. W. J. Conover (1971), Practical Nonparametric Statistics, John Wiley and Sons, Inc., New York, pp. 309-314.

13. R. K. Crane (1976), "An Algorithm to Retrieve Water Vapor Information from Satellite Measurements", NEPRF Technical Report 7-76 (ERT), Concord, MA, November.

14. D. Deirmendjian (1964), "Scattering and Polarization Properties of Water Clouds and Hazes in the Visible and Infrared", Applied Optics, 3, 2, p. 187.

15. W. F. Denham and S. Pines (1966), "Sequential Estimation When Measurement Function Nonlinearity is Comparable to Measurement Error", AIAA Journal, vol. 4, no. 6, pp. 1071-1076.

16. N. Draper and H. Smith (1966), Applied Regression Analysis, John Wiley and Sons, New York.

17. E. Fix and J. L. Hodges (1952), "Discriminatory Analysis, Small Sample Performance", USAF Sch. Aviat. Med., Rep. 11.

18. W. A. Follansbee (1976), "Estimation of Daily Precipitation Over China and the USSR Using Satellite Imagery", NOAA Technical Memo, NESS 81, September.

19. M. G. Fowler, N. D. Sze and N. E. Gaut (1975), "The Estimation of Clear Sky Emission Values from Cloudy Radiometric Data", AFCRL-TR-75-0440, August.

20. M. G. Fowler, K. R. Hardy and N. D. Sze (1976A), "The Development of a Model to Infer Precipitation from Microwave Measurements", ERT Document No. P-1068F, April.

21. M. G. Fowler and A. S. Lisa (1976B), "High Resolution Radiometric Measurements of Convective Storms During the GATE Experiment", ERT Document No. P-1282F, September.

22. M. G. Fowler, J. H. Willand, D.T. Chang and R. G. Isaacs (1977), "Estimation of the Geophysical Properties of the Ocean Surface Using Aircraft Microwave Measurements", ERT Document No. P-1143F, Concord, MA, February.

23. "Frequency Band Justifications for Passive Sensors. 1 to 10 GHz" (1976), NASA, December.

24. S. Fritz and P. K. Rao (1967), "On the Infrared Transmission through Cirrus Clouds and the Transmission of Relative Humidity from Satellites", J. Applied Meteorology, vol. 6, p. 1088, December.

25. F. K. Fye and G. L. Logan (1976), "AFGWC Satellite-Based Analysis and Prediction Programs".

26. F. K. Fye (1978), " The AFGWC Automated Cloud Analysis Model", AFGWC Tech. Memo 78-002.

27. N. E. Gaut, E. C. Reifenstein, III, and D. T. Chang (1972), "Microwave Properties of the Atmosphere, Clouds and the Oceans", Final Report, ERT Project P-280, March.

28. N. E. Gaut, et al. (1974), "Studies of Microwave Remote Sensing of Atmospheric Parameters", AFCRL-TR-75-0007, August.

29. C. G. Griffith and W. L. Woodley (1973), "On the Variation with Height of the Top Brightness of Precipitating Convective Clouds", J. Appl. Meteorology, vol. 12, p. 1086.

30. D. E. Gustafson, et al. (1978), Final Report on Contract F33615-77-C-0606, Scientific Systems, Inc., March.

31. G. R. Hunt, J. W. Salisbury and J. T. Bunting (1974), "Distinction Between Snow and Cloud in DMSP Satellite Imagery: A Preliminary Report", unpublished memo.

32. A. H. Jazwinski (1970), Stochastic Processes and Filtering Theory, Academic Press, New York.

33. R. E. Kalman (1960), " A New Approach to Linear Filtering and Prediction Problems", Trans. ASME, Series D: Jour. Basic Eng., vol. 82, p. 35.

34. B. J. Kilonsky and C. S. Ramage (1976), " A Technique for Estimating Tropical Open-Ocean Rainfall from Satellite Observations", J. Applied Meteorology, vol. 15, p. 972, September.

35. R. Koffler, A. G. DeCotiis and P. K. Rao (1973), "A Procedure for Estimating Cloud Amount and Height from Infrared Radiation Data", Monthly Weather Review, vol. 101, no. 3, p. 240.

36. R. C. Lo and D. R. Johnson (1971), "An Investigation of Cloud Distribution from Satellite Infrared Radiation Data", Monthly Weather Review, vol. 99, no. 8, p. 599.

37. D. O. Loftsgaarden and C. P. Quesenberry (1965), "A Nonparametric Estimate of a Multivariate Density Function", Ann. Math. Statist., vol. 38, pp. 289-290.

38. M. E. Lopez and S. K. Ronning (1964), "A Study of Middle Cloud Wave Patterns", W. E. Howell Associates, Inc., Report on Contract No. AF 19(628)-306.

39. W. S. Meisel and D. A. Michalopoulos (1973), "A Partitioning Algorithm With Application in Pattern Classification and Optimization of Decision Trees", IEEE Trans. on Comp., vol. C-22, pp. 93-103.

40. H. W. O'Brien and R. H. Munis (1974), "Red and Near-Infrared Spectral Reflectance of Snow", USA CRREL Research Report (in press), as cited in Valovcin (1976).

41. E. Parzen (1962), "On the Estimation of a Probability Density Function and the Mode", Ann. Math. Statist., vol. 33, pp. 1065-1076.

42. R. M. Pickett and E. S. Blackman (1976), "Automated Processing of Satellite Imagery Data at Air Force Global Weather Central (AFGWC); Survey, Recommendations and R & D Design Evaluation Report", BBN Report No. 3275, April 23.

43. P. K. Rao (1970), "Estimating Cloud Amount and Height from Satellite Infrared Radiation Data", ESSA Tech. Report, NESC 54, July.

44. M. S. V. Rao and W. V. Abbott (1976), "Quantitative Mapping of Rainfall Rates over the Oceans Utilizing Nimbus-5 ESMR Data", Final Report, Contract No. NAS5-21908, Environmental Research & Technology, Inc., Concord, MA.

45. D. Reynolds and T. H. Von der Haar (1973), "A Comparison of Radar-Determined Cloud Height and Reflected Solar Radiance Measured from the Geosynchronous Satellite ATS-3", J. Applied Meteorology, vol. 12, p. 1082, September.

46. D. W. Reynolds and T. H. Von der Haar (1977), "A Bispectral Method for Cloud Parameter Determination", Monthly Weather Review, vol. 105, pp. 446-457.

47. D. C. Rife, M. Goldstein and R. R. Bourstyn (1975), "A Unification of Cramer-Rao Type Bounds", IEEE Trans. on Information Theory, vol. IT-21, pp. 330-332.

48. D. B. Ross and V. Cardone (1974), "Observations of Oceanic Whitecaps and Their Relation to Remote Measurements of Wind Speed", J. Geophy. Res., 79, 3, pp. 444-452.

49. W. E. Shenk, R. J. Holub and R. A. Neff (1976), "A Multispectral Cloud Type Identification Method Developed for Tropical Ocean Areas with Nimbus-3 MRIR Measurements", Monthly Weather Review, vol. 109, p. 284, March.

50. G. J. Sikula (1974), "Spectral Signatures of Several Cloud Types and Information Extraction from Very High Resolution Visual Satellite Radiances - Preliminary Results", Sixth Conference on Aerospace and Aeronautical Meteorology, El Paso, Texas, November 12-14.

51. W. L. Smith et al. (1972), "The Airborne ITPR Brassboard Experiment", NOAA Technical Report NESS 58, March.

52. W. L. Smith, et al. (1974), "Nimbus-5 Sounder Data Processing System - Part I: Measurement Characteristics and Data Reduction Procedures", NOAA Technical Memo NESS 57, June.

53. D. H. Staelin, et al. (1976), "Remote Sensing of Atmospheric Water Vapor and Liquid Water with the Nimbus-5 Microwave Spectrometer", J. Applied Meteorology, vol. 15, p. 1204, November.

54. "Tuning Factors Applied to Satellite Data in Production of 3DNEPH" (1977), AFGWC Production Division, January 15.

55. "Upper Atmospheric Research Program" (1976), NASA/Goddard Space Flight Center, 2nd Edition, December.

56. S. L. Valley, editor (1965), Handbook of Geophysical and Space Environments, Air Force Cambridge Research Laboratories, Bedford, MA, pp. 7-3 to 7-5.

57. F. R. Valovcin (1976), "Snow/Cloud Discrimination", AFGL-TR-76-0174, August 4.

58. H. L. VanTrees (1968), Detection, Estimation and Modulation Theory, Part I, John Wiley and Sons, New York.

59. W. W. Vickers and B. Thompson (1966), "Optimization of Photo Detail by Holographical Measurements", Tech OPs Report, Burlington, MA, pp. 66.

60. T. J. Wagner (1975), "Nonparametric Estimates of Probability Densities", IEEE Trans. Information Theory, vol. IT-21, pp. 438-440.

61. R. L. Weichel (1976), "Combined Microwave-Infrared Sounding Studies", AFCRL-TR-75-0572, January 12.

62. L. F. Whitney and L. D. Herman (1968), "The Nature of Intermediate-Scale Cloud Spirals", ESSA Tech. Report NESC-45, May.

63. T. T. Wilheit, et al. (1976), "Meteorological Interpretations of the Images from the Nimbus-5 Electrically Scanned Microwave Radiometer", J. Applied Meteorology, vol. 15, p. 166, February.

64. J. H. Willand, M. G. Fowler, E. C. Reifenstein, III and D. T. Chang (1973), "Analysis of Aircraft Microwave Measurements on the Ocean Surface", ERT Document P-442, December.

65. G. Yamamoto, M. Tanaka and S. Asano (1971), "Table of Scattering Function of Infrared Radiation for Water Clouds", NOAA Technical Report NESS 57, April.

APPENDIX A

COMPARISON OF INVERSION METHODS

It is instructive to compare the various inversion techniques explored in this study by means of a simple numerical example. We will consider here the following methods:

(1) linear regression

(2) extended Kalman Filter

(3) iterated extended Kalman Filter

(4) modal estimator

The n-dimensional state x to be estimated is assumed to be normally distributed with mean m and covariance matrix P_o. The p-dimensional measurement y is nonlinear in x:

$$y = h(x) + v \tag{A.1}$$

where v is zero-mean additive noise, normally-distributed with covariance matrix V. The conditional probability density of x, given y, is:

$$p(x|y) = c \exp\left\{ -\frac{1}{2} (x-m)^T P_o^{-1} (x-m) - \frac{1}{2} (y-h(x))^T V^{-1} (y-h(x)) \right\} \tag{A.2}$$

where c is a normalization constant.

The estimators are given as follows:

(1) Linear Regression

$$\hat{x}_L = a + By \tag{A.3}$$

where the $n \times 1$ vector a and the $n \times p$ matrix B are found by minimizing

$$E[(\hat{x}-x)^T (\hat{x}-x)]$$

with respect to a, B and E(·) denotes expectation with respect to the underlying probability density or, in some cases, is an arithmetic average over an ensemble of data.

(2) Extended Kalman Filter

$$x_{EK} = m + K[y-h(m)] \tag{A.4}$$

$$K = P_o H^T [H P_o H^T + V]^{-1} \tag{A.5}$$

where $$H = \frac{\partial h(x)}{\partial x}\Big|_{x=m} \tag{A.6}$$

In addition, the a posteriori covariance is estimated as

$$P_{EK} = P_o - K H P_o \tag{A.7}$$

(3) Iterated Extended Kalman Filter

Starting with the estimate $\beta^o = m$, the following iteration is performed

$$\text{(i)} \quad \beta^{k+1} = m + M^k [y - h(\beta^k) + H^k (\beta^k - m)] \tag{A.8}$$

$$\text{(ii)} \quad M^k = P_o(H^k)^T [H^k P_o (H^k)^T + V]^{-1} \tag{A.9}$$

$$\text{where } H^k = \frac{\partial h(x)}{\partial x}\Big|_{x=\beta^k} \tag{A.10}$$

The iteration terminates when $\|\beta^{k+1} - \beta^k\| < \varepsilon$. If we denote the final value of β as β^q, then

$$\hat{x}_{IEK} = \beta^q \tag{A.11}$$

$$P_{IEK} = P_o - M^q H^q P_o \tag{A.12}$$

(4) Modal Estimator

At the mode of $p(y|x)$ we have, from (A.2):

$$\frac{\partial}{\partial x}\left[(x-m)^T P_o (x-m) + (y-h(x))^T V^{-1} (y-h(x))\right]_{x=\hat{x}_m} = 0 \qquad (A.13)$$

This becomes, after simplifying:

$$\hat{x}_m = m + P_o H_m^T V^{-1} (y-h(\hat{x}_m)) \qquad (A.14)$$

where $$H_m = \frac{\partial h(x)}{\partial x}\bigg|_{x=\hat{x}_m} \qquad (A.15)$$

For nonlinear $h(\cdot)$, (A.14) may not be solvable directly and one may have to rely on iterative methods.

We will now consider the following scalar example:

$$h(x) = x^2, \quad x = 2$$

$$V = 1, \quad V = 1$$

$$m = 1, \quad P_o = 1$$

Thus $y = 5$. The four methods yield the following results.

(1) Linear Regression

$$x_L = a + by$$

where $$b = \frac{2mP_o}{(2P_o + m^2)^2 + V - 2P_o^2}$$

$$= \frac{1}{4}$$

$$a = m - b(P_o + m^2)$$

$$= \frac{1}{2}$$

Thus

$$\hat{x}_L = 1.75$$

(2) Extended Kalman Filter

$$K = \frac{2}{5}$$

$$\hat{x}_{EK} = 1 + \frac{2}{5}(5 - 1)$$

$$= 2.6$$

(3) Iterated Extended Kalman Filter

$$\beta^{(1)} = 1, \qquad M^{(1)} = \frac{2}{5}$$

$$\beta^{(2)} = 1 + \frac{2}{5}\left\{ 5 - 1 + 2(1-1) \right\}$$

$$= 2.6$$

$$M^{(2)} = \frac{2(13)}{5(4(13/5)^2 + 1)}$$

$$= .185$$

$$\beta^{(3)} = 1 + .185\left\{ 5 - (13/5)^2 + (26/5)(13/5 - 1) \right\}$$

$$= 2.2136$$

$$M^{(3)} = .215$$

$$\beta^{(4)} = 2.18$$

.

.

.

(4) Modal Estimator

$$\hat{x}_m = 2.175$$

However, successive approximation won't work here, nor will a gradient method starting with m = 1 as the initial estimate. A gradient method starting from $\hat{x}_{IEK}$ will, however, work. This extreme sensitivity was found to be a problem in multivariate cloud parameter determination from microwave data. Hence, the modal estimator is not recommended as a viable method at this time. Further study is required to find robust methods of finding $\hat{x}_m$ and evaluating its accuracy relative to the other methods.

APPENDIX B

CRAMER-RAO BOUND FOR ESTIMATION ERRORS

The Cramer-Rao bound provides an easily-computed performance limit on the accuracy with which parameters can be estimated from data.

The estimation problem under consideration is the following. The parameters are represented by the $n \times 1$ vector x which is Gaussian distributed with mean m and covariance matrix P_o. Information about x is obtained from a single measurement of the form

$$y = h(x) + v \tag{B.1}$$

where $h(\cdot)$ is a known linear or nonlinear function and $\frac{\partial h(x)}{\partial x}$ exists over the domain of x, Ω_x. The measurement noise v is a zero-mean Gaussian random variable with covariance matrix V, and is independent of x.

The information available to estimate x is contained in y and the parameters m, P_o, V. Thus, we may write the estimate of x as

$$\hat{x} = g(y;m,P_o,V) \tag{B.2}$$

The objective is to find $g(\cdot;\cdot)$ such that some measure of the estimation error, $e = \hat{x} - x$, is minimized. We choose the quadratic form

$$C = E[e^T e] \tag{B.3}$$

where E denotes expectation over all underlying random variables. In practice, we have used ensemble averages over the available data in the data base.

The determination of g(;) to minimize C is straightforward if h(·) is linear and yields the Kalman Filter of Section III. However, difficulties arise if h(·) is nonlinear. In fact, no general solution to the problem exists at the present time and approximations must almost inevitably be made. Rather than search for "good" nonlinear solutions it is more useful in many cases to investigate performance bounds. A particularly useful tool is the Cramer-Rao bound, which can be easily computed in this case, and yields a matrix R such that

$$R \leq E[e\,e^T]^* \tag{B.4}$$

for arbitrary $g(y;m,P_o,V)$. Furthermore, equality holds if and only if h(·) is linear. Notice that this implies the optimality of the Kalman Filter for linear measurements.

We now turn to computation of R.

A.1 General Results

Let p(x|y) be the density function of x, conditioned on y and define

$$H(x|y) = \log p(x|y) \tag{B.5}$$

Let s be an $n \times 1$ vector with elements

$$s_i = \frac{\partial H}{\partial x_i} \tag{B.6}$$

Then the Fisher information matrix is

$$J = E[s\,s^T] \tag{B.7}$$

---

*The matrix inequality $A \leq B$ means that $u^T A\,u \leq u^T B\,u$ for all real u.

Then, following VanTrees (1969), the lower bound is of the form

$$R = J^{-1} \tag{B.8}$$

if the estimate is unbiased; i.e.

$$E[g(y;m,P_o,V) - x] = 0 \tag{B.9}$$

which is a desirable property of an estimator in practice.

A.2 Evaluation of J for Gaussian Random Variables

For x, v Gaussian, we have

$$\log p(x|y) = \text{const} - \frac{1}{2}(x-m)^T P_o^{-1}(x-m) - \frac{1}{2}(y-h(x))^T V^{-1}(y-h(x)) \tag{B.10}$$

from which

$$s = h_x^T V^{-1}(y - h(x)) - P_o^{-1}(x-m) \tag{B.11}$$

where $h_x = \partial h/\partial x$ and $[\partial h/\partial x]_{ij} = \partial h_i(x)/\partial x_j$

Thus, setting $\Delta x = x - m$ and using $y - h(x) = v$, we have

$$J = E\left[h_x^T V^{-1} v - P_o^{-1}\Delta x\right]\left[v^T V^{-1} h_x - \Delta x^T P_o^{-1}\right] \tag{B.12}$$

Using $E(\Delta x \Delta x^T) = P_o$

$E(v\,v^T) = V$

$E(v\,\Delta x^T) = 0$

in (B.12) gives

$$J = P_o^{-1} + E(h_x^T V^{-1} h_x)$$

Thus, for any unbiased estimator

$$\hat{x} = g(y;m,P_o,V)$$

$$E[(\hat{x}-x)(\hat{x}-x)^T] \geq J^{-1}$$

which is the desired result.

Note, in particular, that the variance of error for the ith parameter is bounded by

$$\text{var}(x_i) = E[(\hat{x}_i - x_i)^2]$$

$$\geq [J^{-1}]_{ii}$$

This result is used in Section III to provide performance bounds using different data channels and different noise magnitudes.